

# Encoding of natural sounds in the human brain

Citation for published version (APA):

Moerel, M. M. L. (2013). *Encoding of natural sounds in the human brain*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20130328mm>

**Document status and date:**

Published: 01/01/2013

**DOI:**

[10.26481/dis.20130328mm](https://doi.org/10.26481/dis.20130328mm)

**Document Version:**

Publisher's PDF, also known as Version of record

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

## **Encoding of natural sounds in the human brain**

© M.M.L. Moerel, Maastricht 2013

Cover	deviantART
Production	CPI Wöhrmann Print Service B.V.
ISBN	978-94-6203-294-1

The work presented in this thesis was funded by the Toptalent program of the Netherlands Organization for Scientific Research (NWO) and was conducted at Maastricht University.

# **Encoding of natural sounds in the human brain**

Dissertation

To obtain the degree of Doctor at the Maastricht University,  
on the authority of Rector Magnificus, Prof. dr. L.L.G. Soete,  
in accordance with the decision of the Board of Deans,  
to be defended in public on Thursday 28<sup>th</sup> of March at 10.00 hours

by

Michelle Maria Lucia Moerel

**Supervisor**

Prof. dr. E. Formisano

**Co-supervisor**

Prof. dr. R. Goebel

Dr. F. De Martino

**Assessment Committee**

Prof. dr. P. De Weerd (Chairman)

Prof. dr. P. Belin (University of Glasgow, UK)

Prof. dr. M. Sereno (University College London, UK)

Dr. K. Uludag

Dr. M. Bonte

# Contents

Chapter 1	General introduction	07
Chapter 2	Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning and relation to voice-sensitivity	31
Chapter 3	Spatial organization of frequency preference and selectivity in the human inferior colliculus	61
Chapter 4	Processing of natural sounds: Characterization of multi- peak spectral tuning in human auditory cortex	85
Chapter 5	The perceptual constancy of musical notes across octaves emerges from multi-peak spectral tuning in primary auditory cortex	111
Chapter 6	General discussion	133
	Summary	145
	Samenvatting	147
	Acknowledgements	149
	Curriculum vitae	151
	Publications	153



# Chapter 1

## General introduction





Listening to our favourite piece of music, navigating through traffic, or interacting with colleagues: we are constantly surrounded by sounds. These sounds are nothing but pressure waves reaching our ears; yet they elicit perception, emotion, and action. Music might give us joy, the sound of a car horn makes us jump quickly back onto the sidewalk, and a conversation with a colleague may enhance our understanding of a difficult topic. These examples illustrate the richness of the information that the human brain can extract from sound waves arriving at the ear.

After transduction in the cochlea, incoming sounds are analyzed and transformed into behaviourally useful neural representations within the various structures that compose the auditory system. Starting at the auditory periphery, the various acoustic features of sounds (e.g., frequency, intensity, spectral and temporal modulations, and binaural differences) are processed. Auditory neurons often show selective responses to these features, and - for example - can be characterized according to their preferred frequency, intensity, and modulation rate (Merzenich et al., 1973; Philips and Irvine, 1981; Joris et al., 2004). This suggests that dedicated neural mechanisms exist in the auditory system for processing the acoustic features of natural sounds.

This thesis addresses two largely unanswered questions regarding auditory feature processing. First, a wealth of information regarding feature processing mechanisms was established by recording brain responses to simple synthetic sounds, mostly in anesthetized animals. Are the mechanisms unravelled in these studies also used when awake humans process natural sounds? Second, imaging studies in humans have indicated that at higher levels of the auditory hierarchy an abstract, categorical sound representation is formed (e.g. Belin et al., 2000; Leaver et al., 2010). How are low-level features combined to ultimately create a higher-level categorical sound representation?

Understanding the tuning properties of the human auditory system and the mechanisms by which a higher order sound representation is created, is crucial for advancing the knowledge regarding the human auditory system in general. Moreover, this knowledge may be relevant for improving our understanding of auditory malfunctions. The experimental studies forming this thesis aim at advancing this knowledge. The present chapter discusses research findings that served as background for these studies. Specifically, acoustic properties of natural sounds are introduced, and the anatomical and functional properties of subcortical and cortical auditory structures are outlined (for an exhaustive review, see Winer and Schreiner, 2005; 2010). Furthermore, an introduction to the employed signal acquisition and analysis methods is given.

## Acoustic features of natural sounds

Natural sounds are the sounds that naturally occur in everyday life<sup>1</sup>. As you pay attention to the sounds currently surrounding you, there are various factors by which you can discriminate one from the other. One sound may be louder than the other, their rhythm may be different, or one of the sounds may be higher in pitch. These observations describe perceptual correlates of the sound features amplitude, temporal modulations, and frequency, which are examples of the building blocks of natural sounds. While some features may be easy to assign to the sounds surrounding you, many features can only be defined based on a formal computational analysis of the sounds. Much work has been done to characterize natural sounds in terms of sound features, and accordingly a large number of acoustic features have been described (Singh and Theunissen, 2003; Cavaco and Lewicki, 2007; Gygi et al., 2007). However, to date the understanding of which low-level features are relevant for acoustic analysis in the brain is still limited (King and Nelken 2009; Recanzone and Cohen 2010).

An intrinsic relationship may exist between acoustic features and certain natural sound categories. For example, human speech has a relatively low spectral content, slow temporal fluctuation, and a characteristic harmonic spectral profile (Crandall and Mackenzie, 1922; Lewis et al., 2009). The harmonicity of speech sounds stems from the vibrations of the vocal folds in the larynx as air flows through them, resulting in periodic sounds. Harmonic complexes are characteristic for many natural sounds, for example for animal calls and the sounds of musical instruments. In contrast, environmental scenes (e.g. sounds of the forest) cover a very wide spectral range and are noisy (no harmonic components; Singh and Theunissen, 2003). The intrinsic relation between acoustic features and sound categories complicates investigations of brain processes related to feature- and categorical sound representations (Staeren et al., 2009). For example, an observed neuronal population preference for vocalizations may indicate preference for features common to vocalizations instead of a categorical representation (Leaver et al., 2009). This issue is further discussed in chapter 2.

---

<sup>1</sup> As this definition is not unequivocal, I will use the term 'natural sounds' in this thesis to include human and animal vocalizations, musical instruments, man-made tool sounds, and environmental scenes.

## Basic principles of the neuronal sound processing

After sound waves enter the ears, they are processed by a hierarchically ordered system of auditory relays. Below, this hierarchy is discussed in three parts (i.e. the periphery, subcortical processing, and the cortex), and a brief introduction to the anatomical and functional properties of each part is given. For a detailed overview of basic sound processing, see Bear et al. (2007).

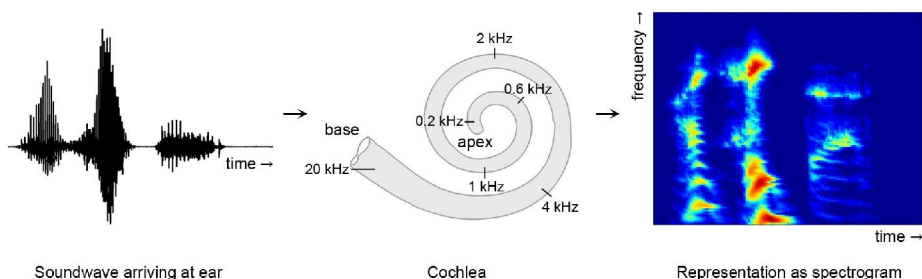
### The auditory periphery

The auditory periphery includes processing stages from the outer ear until the cochlear nerve. Although the distance between these structures is small, a crucial sound transformation is performed within these early stages. This transformation is often compared to a Fourier analysis of the incoming sound wave, and it defines how sounds are processed at the later stages of the auditory hierarchy. Sounds enter the outer ear as pressure waves (left side of Figure 1), which are transmitted through the middle ear to produce vibrations at the oval window. These vibrations are transferred to the cochlear fluid, where they create a pressure wave that travels down the cochlea located in the inner ear. Two membranes, the basilar membrane (BM) and Reissners membrane, divide the cochlea into three chambers. Because stiffness of the BM changes throughout the cochlea, the displacement of the membrane induced by an incoming sound depends on the frequency of that sound. Specifically, the BM is stiffer near the oval window (the ‘base’) than near the middle of the spiral (the ‘apex’). Consequently, high frequency sounds produce displacement near the base, while low frequency sounds disturb the membrane near the apex (see Figure 1).

As a result, each BM site is identified by its characteristic frequency (CF), and the whole BM can be described as a bank of overlapping filters (e.g. Patterson et al., 1992; Meddis and Lopez-Poveda, 2010). The organ of Corti resides on the BM and contains the inner hair cells (IHC), which are responsible for the mechano-electrical transduction. As the BM displaces, it causes deflection in the stereocilia of the location-matched IHC, which results in a current flow and ultimately an action potential. Because BM displaces in a frequency dependent manner, the corresponding IHC are “tuned” to sound frequency. The resulting spatial proximity of contiguous preferred sound frequency is referred to as tonotopy (or cochleotopy). This tonotopic organization is carried up through the auditory hierarchy to the cortex, and defines the functional topography in each of the intermediate relays.

The computations performed by the auditory periphery are more complex than the simple Fourier analogy given here, and include various nonlinear mecha-

nisms that optimize the sound representation for subsequent processing (Meddis and Lopez-Poveda, 2010). These stages have been investigated extensively, and are well characterized by computational models (Chi et al., 2005; Balaguer-Ballester et al., 2008). In the current thesis, the computational model proposed by Chi et al. (2005) is used in each chapter to describe the representation of natural sounds after early sound processing stages. Subsequently, this sound representation is used as input to the encoding model of cortical processing (example of model output in spectrogram on the right side of Figure 1).



**Figure 1. Cochlear mechanisms represent sounds as spectrograms**

Processing in the auditory periphery can be compared to performing a Fourier analysis on the incoming sound wave. As a sound wave enters the ear (here the example is a 1 sec stimulus of a male speaker saying “I hate that”; left), it is transmitted to the cochlea where it displaces the basilar membrane (BM) in a frequency dependent manner (middle). As a result, the whole BM can be described as a bank of overlapping filters, which transform the sound representation to that of a spectrogram (frequencies changing over time; right).

### Subcortical auditory processing

After peripheral processing, the acoustic spectrogram is encoded in the complex pattern of action potentials and neuronal activity that travels through the spiral ganglion cells of the auditory nerve (cranial nerve VIII) to the cochlear nucleus (CN; Figure 2). In the CN, specialized cell types sharpen frequency and time information (Schneggenburger and Forsythe, 2006; Golding and Oertel, 2012). The CN sends information to the ipsilateral and contralateral regions of the superior olivary complex (SOC), which computes differences in time and intensity of sounds arriving at the two ears (interaural time and level differences; Oliver et al., 2003; Tollin et al., 2008). Consequently, the SOC plays a prominent role in sound localization. From the SOC, and in addition from the CN, the neural signals are carried to the central nucleus of the inferior colliculus (IC) through a dense fiber bundle called the lateral lemniscus.

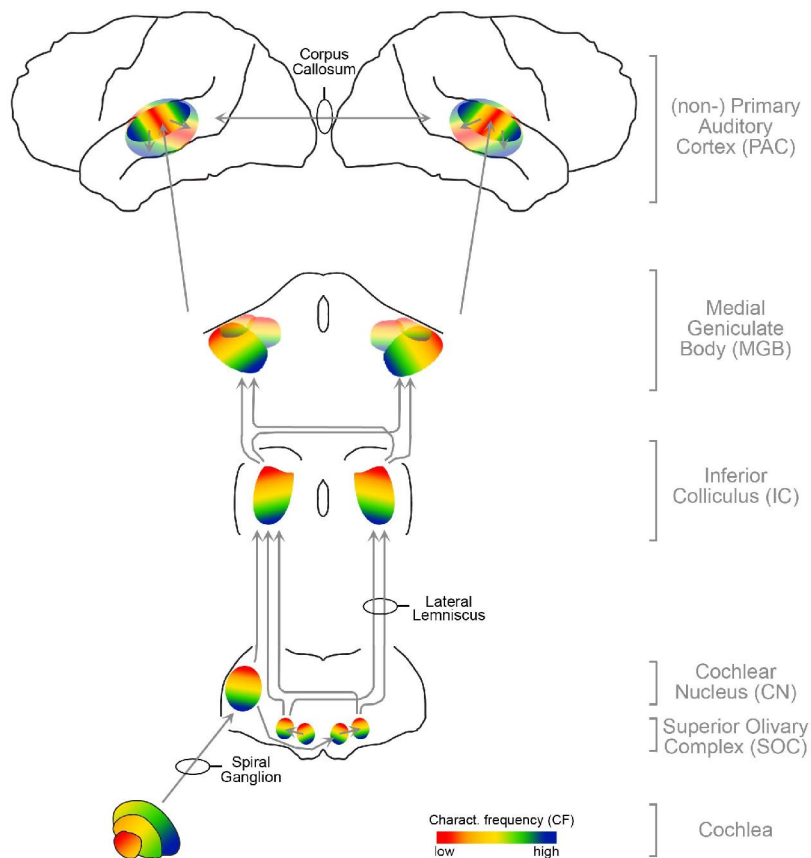
The IC is an obligatory relay for all ascending auditory information (Figure 2), and as such integrates output from CN and SOC regions before sending it on to the thalamus. The IC is a laminar structure, and the main tonotopic gradi-

ent – which is carried up from the cochlea by a rigid arrangement of ascending fibers – runs orthogonal to the fibrodendritic laminae in the central nucleus of the IC in dorso-lateral to ventro-medial direction (Merzenich and Reid, 1974, see chapter three). That is, isofrequency bands coincide with a lamina, creating smooth frequency changes within a lamina and systematic frequency discontinuities across lamina (Schreiner and Langer, 1988; Malmierca et al., 2008). These step-wise changes of approximately  $\frac{1}{3}$  octave in best frequency have been related to perceptual phenomenon of critical bandwidth (i.e. the psycho-physical spectral range within which humans integrate frequencies; Schreiner and Langer, 1997). Beyond frequency, other acoustic features such as spectral and temporal modulations may be topographically represented within the IC (Rodriguez et al., 2010; Baumann et al., 2011).

The IC sends its output to the medial geniculate body (MGB) of the thalamus (Wenstrup, 2005). The MGB consists of three major divisions, of which at least one (ventral MGB) and perhaps all three contain a full tonotopic maps (see Figure 2). The ventral division of the MGB projects directly to the first cortical processing stage: the primary auditory cortex (Hackett et al., 2011).

## The auditory cortex

The human auditory cortex is situated on the supratemporal plane, and comprises the superior two-thirds of the superior temporal gyrus (STG; Celesia 1976; Galaburda and Sanides, 1980; Rivier and Clarke, 1997; Figure 3). For the monkey, a generally accepted hierarchical model of auditory cortex exists, consisting of a “core” auditory cortex and surrounding belt and parabelt regions (Rauschecker et al., 1995; Hackett et al., 2001; Rauschecker and Tian 2004; Figure 4A). Although several studies indicate that human auditory cortex contains a similar hierarchy (Wessinger 2000, Chevillet et al., 2011), a complete model of human auditory cortex is still lacking. That is, the number, organization and functional role of the auditory regions covering the human superior temporal lobe are still debated (Humphries et al., 2010; Da Costa et al., 2011; Moerel et al., 2012). Below we review macro-anatomy, microanatomy, and functional properties of the human auditory cortex. We compare them to the monkey auditory model for the purpose of interpretation.

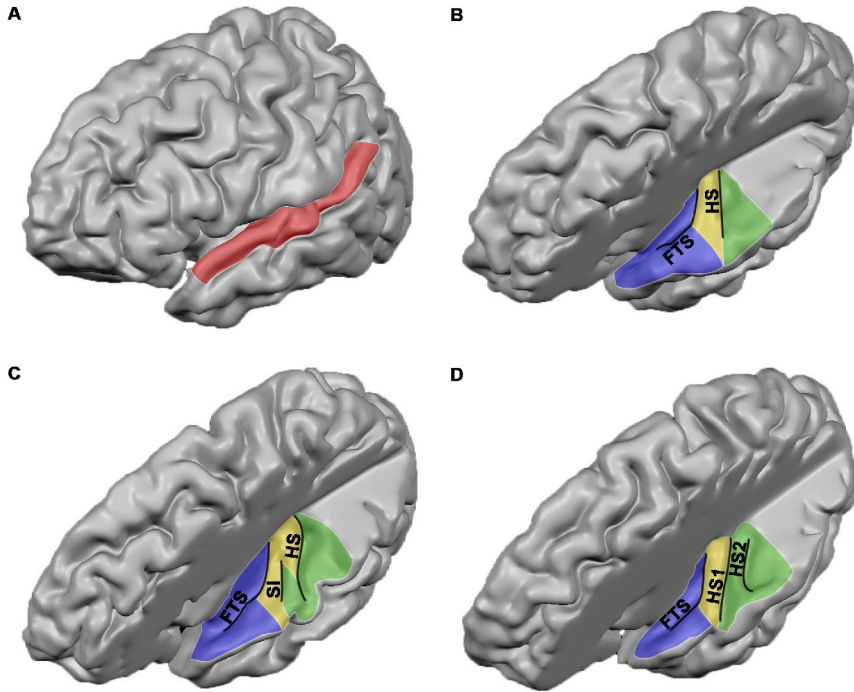


**Figure 2. Frequency preference in ascending auditory pathway**

Major structures of the ascending auditory pathway are shown, with their frequency preference. Incoming sounds displace the cochlea dependent on their frequency content, and from that point on in the auditory hierarchy, sound frequency bands are selectively processed in spatially segregated manner. Auditory neurons with similar frequency preference cluster together, creating one or multiple tonotopic maps per auditory level.

## Macro-anatomy

On a macroscopic scale, the human auditory cortex can be divided in three regions (Kim et al., 2000; Figure 3). In anterior to posterior direction, the auditory cortex includes planum polare (PP), the transverse temporal gyrus or Heschl's gyrus (HG), and planum temporale (PT). HG is a convolution on the supratemporal plane, branching obliquely from the STG and hidden in the depth of the Sylvian fissure (SF). HG is evolutionary new: this convolution is not



**Figure 3. Anatomical landmarks on the human supratemporal plane**

(A) Lateral view of left hemisphere, with STG indicated in red. (B-D) Top view of left supratemporal plane, after removal of a large part of the parietal cortex. PP, HG, and PT are indicated in blue, yellow, and green respectively. Major sulci are outlined in black (first transverse sulcus [FTS]; sulcus intermediate [SI]; Heschl's sulcus [HS]; first Heschl's sulcus [HS1]; second Heschl's sulcus [HS2]). Panels include hemispheres with one HG, an incomplete separation of HG, and two HG in B-D respectively.

present in the macaque monkey, and can be discerned in only a subset of chimpanzee brains (Hackett et al., 2001). There is considerable variability in the number of convolutions on the human supratemporal plane, ranging from one to three complete duplications of the transverse gyrus per hemisphere (compare Figures 3B to 3D; Campain and Minckler, 1976; Penhune et al., 1996). Besides complete duplications, a shallow intermediate sulcus (SI) may divide a single HG incompletely (Figure 3C). Based on results from cytoarchitecture (Rademacher et al., 1993), HG is taken as the first (i.e. most anterior) gyrus. HG is bordered medially by the insular cortex, laterally by STG, and anterior and posterior by the first transverse sulcus and Heschl's sulcus respectively (but see variations in Figure 3B-D). PT is located posterior to HG. This triangular region is limited medially by the Sylvian fissure, and laterally by the rim of the su-



pratemporal plane. It shows a marked asymmetry and is consistently larger in the left hemisphere (Geschwind and Levitsky, 1968; Galaburda et al., 1978). In humans, the PT region is much expanded compared to the monkey (Galaburda et al., 1978). Anterior to HG - separated by the FTS - lays PP, further delimited by the insula and the frontal operculum (Kim et al., 2000).

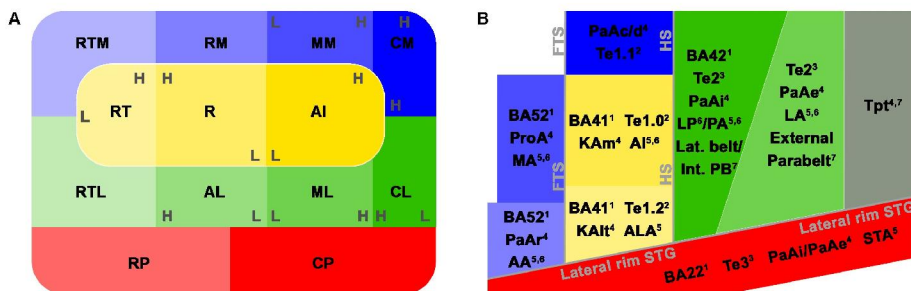
### Microanatomy

In addition to describing the human auditory cortex in terms of its major anatomical landmarks, it has been labelled according to a variety of architectonic schema (Galaburda and Sanides, 1980; Rivier and Clarke, 1997; Hackett et al., 2001; Morosan et al., 2001; Sweet et al., 2005). Across cytoarchitectonic studies, however, large differences exist with respect to the number of observed auditory areas, the location of these regions, and nomenclature. These differences already exist when parcellating HG, yet discrepancies between studies enlarge with increased distance from HG. Here, I attempt to give an overview of obtained results and propose how the different studies may be reconciled (see Figure 4B and Table 1).

All cytoarchitectonic studies delineate homologues region to monkey core or primary auditory cortex (PAC), referring to highly granular and heavily myelinated koniocortex within the auditory cortex. Brodmann (1909) named this area BA 41, and Rivier and Clarke (1997) referred to it as AI. In accordance with the monkey auditory core, which includes two (auditory area 1 [A1], rostral field [R]) or three (A1, R, and rostrotemporal field [RT]) subdivisions (Rauschecker et al., 1995; Hackett et al., 2001), several studies divided the human core into subfields (see Figure 4), most likely reflecting the human homologs of monkey A1 and R (KAM and KAlt: Galaburda and Sanides, 1980; Te1.1, Te1.0, and Te1.2: Morosan et al., 2001), and possibly RT (PaAr, Galaburda and Sanides, 1980; ALA, Wallace et al., 2002). The position of the human PAC relative to sulcal and gyral landmarks is variable. When only one HG is present, the PAC occupies its medial two-third portion. However, when other combinations of HG are present (occurring in the majority of the population), the PAC may extend postero-medially into HS and even onto the second HG (Galaburda and Sanides, 1980; Rivier and Clarke, 1997; Hackett et al., 2001; Morosan et al., 2001; Sweet et al., 2005). Thus, to date a cytoarchitectonic parcelation is needed to unequivocally determine the location of the PAC in humans.

In monkey auditory cortex, a belt region encompassing 7 or 8 subfields surrounds the core. The lateral belt contains regions CL, CM, and AL (Figure 4A; Hackett et al., 1998; 2001; Rauschecker & Tian, 2004), of which field CM seems to be intermediate in hierarchy between core and belt region (Hackett et

al., 2001). Immediately adjacent to the lateral belt on the exposed surface of the STG lay rostral and caudal parabelt (see Figure 4A; Hackett et al., 2001). In accordance with the monkey auditory cortex, in the human cortex several less granular fields surround the PAC. Occupying HS - lateral and immediately adjacent to the PAC – an area with a reduced granular structure compared to primary core areas (parakoniocortex) and with large pyramidal neurons in layer IIIc has been consistently reported (PaAi, Galaburda and Sanides, 1980; Te2, Morosan et al., 2001; 2005; LP, Wallace et al., 2002; lateral belt, Sweet et al., 2005). Posterior to HS, bordering PaAi and extending along the STG, Galaburda and Sanides (1980) distinguished an additional region named PaAe. This region may correspond to posterior BA 42. At this approximate cortical region, other studies described subfields oriented in medial to lateral direction (areas PA, LA, and STA; Rivier and Clarke, 1997; Wallace et al., 2002) but also in caudorostral direction (internal and external parabelt; Sweet et al., 2005; see Table 1 and Figure 4). Posterior to this latter region extending towards the temporoparietal junction, area Tpt is located. Tpt extends beyond the PT, including the postero-lateral STG, portions of the parietal operculum, and part of the supramarginal gyrus (Galaburda and Sanides, 1980; Sweet et al., 2005). Cytoarchitecturally, this is a transitional region between specialized sensory and more general cortex.



**Figure 4. Model of primate and human auditory**

The colors indicate macro-anatomical structures: medial belt/PP (blue), core/HG (yellow), lateral belt/PT (green), and parabelt/STG (red) for monkey and human cortex respectively. (A) Each part of monkey auditory cortex is divided in several auditory fields (auditory area I [AI], rostral [R] and rostrotemporal [RT] area, caudomedial [CM], middle medial [MM], rostromedial [RM], medial rostrotemporal [RTM], caudal lateral [CL], middle lateral [ML], anterior lateral [AL] and lateral rostrotemporal [RTL] belt and caudal and rostral parabelt [CP and RP]; Hackett et al., 1998; Hackett et al., 2001; Rauschecker and Tian, 2004). Neuronal preference for low ('L') and high ('H') frequencies is indicated (adapted from Hackett et al., 1998). (B) Major landmarks are outlined and named in gray (FTS: first transverse sulcus; HS: Heschl's sulcus; STG: superior temporal gyrus). The layout and combinations are interpreted from Brodmann (1909)<sup>1</sup>, Morosan et al. (2001)<sup>2</sup>, Morosan et al. (2005)<sup>3</sup>, Galaburda and Sanides (1980)<sup>4</sup>, Wallace et al. (2002)<sup>5</sup>, Rivier and Clarke (1997)<sup>6</sup>, and Sweet et al. (2005)<sup>7</sup>.

Additional non-primary auditory areas have been reported at the anteromedial portion of the superior temporal plane, extending from the insula to the parietal operculum (area PaAc/d in Galaburda and Sanides, 1980), antero-medially to the PAC, at the border between insular and temporal cortex (BA 52 in Brodmann, 1909; area ProA in Galaburda and Sanides, 1980; area MA in Wallace et al., 2002) and on the planum polare (area AA in Wallace et al., 2002). The description of correspondences between monkey and human auditory cortex beyond the PAC is complicated by evolutionary recent expanses in the human cortex (Galaburda et al., 1978), and the lack of thorough knowledge of functional properties of these auditory cortical areas in either monkey or human (Schreiner and Winer, 2007). Comparative studies are needed to further our understanding of homologies and differences in the functional neuroanatomy of the human and monkey auditory cortex. To this end, large progress is expected from the acquisition of reliably functional MRI (fMRI) measurements in the monkey brain (Petkov et al., 2006; Remedios et al., 2009; Joly et al., 2011).

Brodmann <sup>1</sup>	Morosan <sup>2,3</sup>	Galaburda <sup>4</sup>	Wallace <sup>5</sup>	Rivier <sup>6</sup>	Sweet <sup>7</sup>	Hackett <sup>8</sup>
41	Te1.0	KAm	AI	AI	HG	AI
41	Te1.2	KAlt	AI	AI	HG	R
41	-	PaAr	ALA	-	-	RT
42	Te1.1	PaAc/d	-	-	-	CM/CL
42	Te2	PaAi	LP	-	Lt Belt	ML/AL
42	Te2	PaAe	LA/PA	LA/PA	I/E Pb	-
22	Te3	Tpt	-	-	Tpt	-
22	Te3	PaAe	STA	STA	-	Parabelt
52	TI1	-	AA	AA	-	RTM
52	TI1	ProA	MA	MA	-	RM

**Table 1. Comparison of human cytoarchitectonics and primate fields**

Interpreted from Brodmann (1909)<sup>1</sup>, Morosan et al. (2001; 2005)<sup>2,3</sup>, Galaburda and Sanides (1980)<sup>4</sup>, Wallace et al. (2002)<sup>5</sup>, Rivier and Clarke (1997)<sup>6</sup>, Sweet et al. (2005)<sup>7</sup>, and Hackett et al. (2001)<sup>8</sup>.

## Cortical tuning to frequency

Numerous studies have used electrophysiological recordings to investigate tonotopy in the auditory cortex. Although tonotopy recently has been shown to

break down at the level of individual neurons (Bandyopadhyay et al., 2010; Rothschild et al., 2010), at a larger spatial scale tonotopic maps can reliably be found in the auditory cortex (Merzenich et al., 1973; Reale & Imig, 1980). In primates, tonotopic maps are present in the core auditory region (Merzenich & Brugge, 1973), with reversals in the frequency gradient indicating the borders between the separate auditory fields (AI, R, and RT; see Figure 4A). The low-frequency border shared between A1 and R and high-frequency border between R and RT appear to coincide with histologically defined borders (Merzenich and Brugge, 1973; Morel et al., 1993; Kaas and Hackett, 2000). The frequency selectivity or tuning width – reflecting the range of frequencies to which a neuron responds – is narrow in core regions (Rauschecker et al., 1995; Hackett et al., 1998; Rauschecker & Tian, 2004; Kajikawa et al., 2005; Kusmirek and Rauschecker, 2009).

In the belt areas, a number of auditory fields (e.g. anterolateral field [AL], middle lateral field [ML], caudolateral field [CL], caudomedial field [CM], and middle medial field [MM]) have also been shown to contain a tonotopic map (Merzenich & Brugge, 1973; Rauschecker et al., 1995; Kosaki et al., 1997; Rauschecker & Tian, 2004; Kusmirek & Rauschecker, 2009). The primary frequency gradient (in the regions R, AI, and CM) runs parallel to the gradient in belt areas (AL, ML, and CL, respectively; Rauschecker & Tian, 2004; see Figure 4A). Consequently, reversals in the tonotopic gradient, used to divide core and belt into subfields, cannot be used to distinguish core from belt auditory cortex. Commonly, tuning width is used to achieve this feat, as belt regions have a broader tuning than core areas (Rauschecker et al., 1995; Hackett et al., 1998; Rauschecker and Tian, 2004; Kajikawa et al., 2005; Kusmirek and Rauschecker, 2009).

fMRI studies in humans have partially confirmed the functional organization of the monkey auditory system. Early studies (Bilecen et al., 1998; Talavage et al., 2000; Engelien et al., 2002; Schönwiesner et al., 2002) gathered evidence for the presence of multiple frequency-selective responses along the Heschl's region, but failed to obtain detailed topographical representations of these frequency-selective responses. In one of the first neuroscientific applications of ultra-high field MR (7 Tesla), Formisano et al. (2003) depicted the detailed tonotopic layout of human PAC. Based on the spatial arrangement and mirror-symmetry of the frequency-selective responses, this tonotopic map was interpreted as reflecting the human homologues of monkey areas A1 and R (hA1 and hR; Merzenich et al., 1973; Merzenich & Brugge, 1973; Reale & Imig, 1980; Kaas & Hackett, 2000).

In recent years, the extraction of tonotopic maps throughout the human superior temporal plane with fMRI has become increasingly feasible (Talavage

et al., 2004; Woods et al., 2009; 2010; Humphries et al., 2010; Da Costa et al., 2011; Striem-Amit et al., 2011; Langers and van Dijk, 2012). Resulting maps show great correspondence across studies. A large low frequency region on HG is consistently observed, surrounded posteriorly (on HS and PT), antero-medially, and antero-laterally (on PP) by regions preferring high frequencies. As the primary frequency gradient is assumed to run parallel to the gradient in belt areas, the extraction of tonotopic maps alone is insufficient for determining borders between core and surrounding cortex. Consequently, the interpretation of the observed cortical tonotopy maps is not agreed upon. Some studies interpret the complete high-low-high map, stretching from PP to PT, as reflecting two primary auditory fields hA1 and hR (Da Costa et al., 2011; Langers and van Dijk, 2012). Other studies suggest that part of this large gradient reflects auditory belt fields (Talavage et al., 2004; Woods et al., 2009; 2010; Humphries et al., 2010; Striem-Amit et al., 2011), yet they disagree on which parts of the gradient should be assigned to core and belt cortex respectively. To date, the exact orientation of the main frequency gradient, and therefore of the human PAC, remains issue of debate (Humphries et al., 2010; Da Costa et al., 2011; Langers and van Dijk, 2012).

### **Higher level auditory representations**

Beyond a tonotopic organization, imaging studies revealed that human auditory cortex contains functionally specialized processing streams and regions, similar to the architecture of the visual cortex (Ungerleider & Mishkin, 1982; Kanwisher et al., 1997). There is ample evidence for a ventral sound recognition (“what”) stream and dorsal sound localization (“where”) stream, in which a hierarchy of processing stages is assumed to support the emergence of a more complex and behaviourally relevant sound representation (Tian et al., 2001; Rauschecker & Scott, 2009). Furthermore, studies revealed localized cortical regions in which responses to a preferred category (e.g. voices, music) are stronger than responses to various control sounds (Belin et al., 2000; Zatorre et al, 2002; Leaver & Rauschecker, 2010). These regional activations are thought to be the expression of “categorical” sound representations for which the relation to the acoustic make-up is partially lost (Belin et al., 2000; Zatorre et al., 2002; Leaver and Rauschecker, 2010). Conversely, recent studies showed that sound category information could be detected with multivoxel pattern analyses (Haxby et al., 2001) of spatially distributed activation patterns over the auditory cortices, including areas adjacent or along the HG. These observations suggested a relevant role for non-hierarchical and distributed mechanisms in the processing of complex stimulus attributes and in the formation of perceptual representa-

tions (Formisano et al., 2008; Staeren et al., 2009; Killian-Huetten et al., 2011). Whether hierarchical or distributed, to date the computational mechanisms that underlie the transition from a tonotopic, acoustically faithful sound image into these more abstract, behaviourally relevant and categorical sound representations in the human auditory cortex remain largely elusive.

## **Investigating sound representations in the human brain**

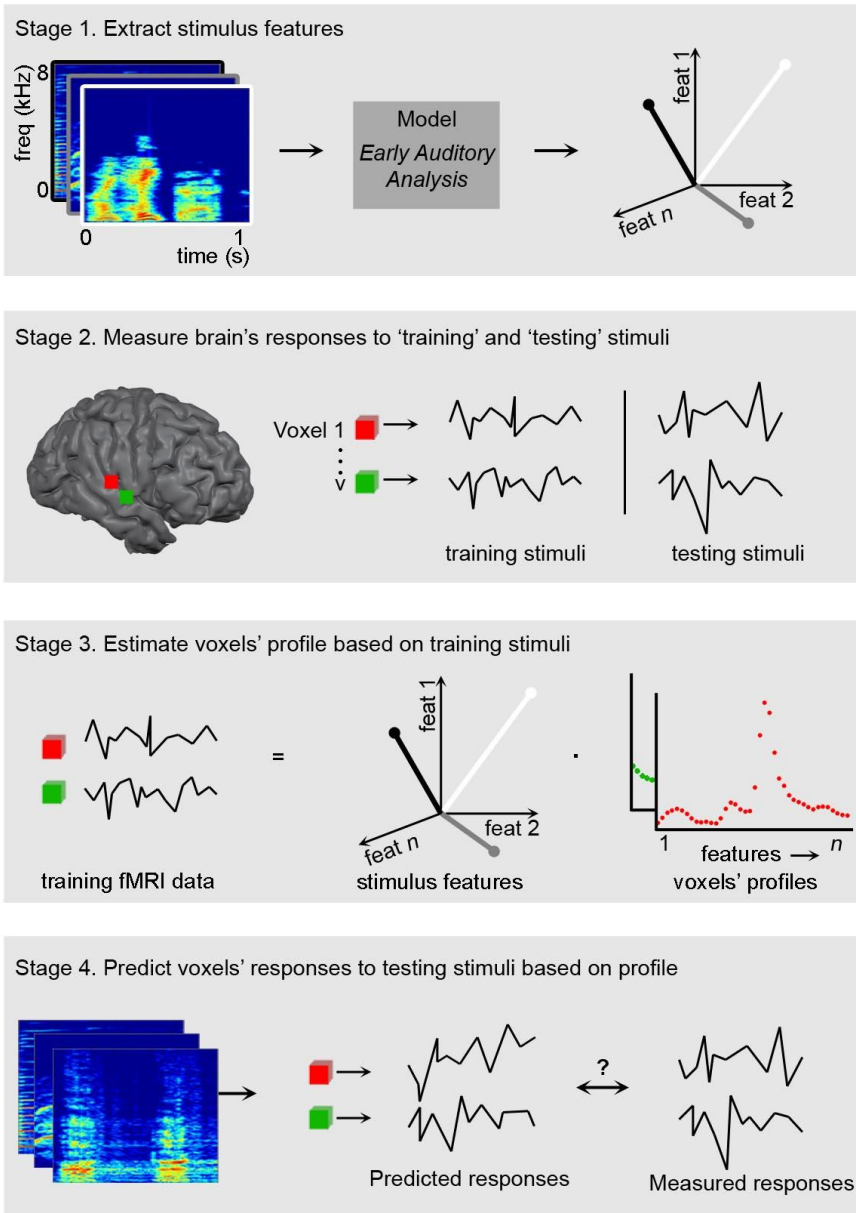
In the present thesis, sound representations throughout the human brain are investigated non-invasively with functional magnetic resonance imaging (fMRI). fMRI is combined with an encoding approach (Kay et al., 2008; Naselaris et al., 2010), with the specific aim of deriving an fMRI-based description of neuronal populations (i.e. voxels) in terms of their feature tuning based on responses to natural stimuli. Using fMRI, neuronal activity is measured indirectly at a spatial scale of a few millimetres (i.e. a voxel). Each voxel includes thousands of neurons, and correspondingly the tuning profile estimated with the encoding approach in each voxel reflects the combined tuning of large neuronal populations. This is a relevant distinction with electrophysiological studies in animals, where similar methods are used for estimating the tuning profiles of single neurons or small neuronal populations.

### **Functional magnetic resonance imaging**

When external stimuli or cognitive processes activate neurons, these neurons increase their oxygen consumption. Oxygen is extracted from the surrounding blood vessels, and after an initial dip in blood oxygenation, blood flow and correspondingly blood volume are increased. Thereby, the ratio between oxygenated and deoxygenated haemoglobin at the site of activation increases, reaching a peak approximately 5 seconds after the event. As oxygenated and deoxygenated haemoglobin have different magnetic properties, these stimulus-induced changes can be measured with fMRI. The change in MR signal triggered by neuronal activity is known as the blood-oxygenated-level dependent (BOLD) hemodynamic response (Ogawa et al., 1992). Note that by using the BOLD contrast, neuronal firing is measured indirectly through corresponding changes in blood oxygenation (Goense and Logothetis, 2008). Thus, although the spatial resolution of fMRI is superior to any other non-invasive imaging method and rapidly improves with recent developments (e.g. ultra-high field imaging, see below), it is ultimately bound by the resolution of the underlying vasculature.

Compared to conventional field strengths (3 Tesla and below), ultra-high magnetic fields (7 Tesla and above) have several crucial advantages. The BOLD contrast increases supra-linearly with magnetic field strength (Yacoub et al., 2001), which together with increases in signal-to-noise ratios (Vaughan et al., 2001) permits increasing the measurements spatial resolution (i.e. decrease the voxel size). Additionally, functional neuroimaging measured at ultra-high fields benefits from increased specificity represented by a reduced point spread function in conventional gradient-echo images (Yacoub et al. 2001; Koopmans et al. 2010; Polimeni et al. 2010), and consequently functional measures at high fields reflect more closely the underlying neuronal population activity. Ultra-high field measurements are used in chapters 3, 4, and 5 of this thesis.

As fMRI data is acquired, rapid switching of the gradient coils during the spatial encoding of the MR signal generates loud acoustic noise (~100 dB sound pressure level; Di Salle et al., 2003). This represents an additional challenge for studying audition, as the noise activates the auditory cortex and interferes with the experimental task of interest. To deal with this noise, passive attenuation is used (ear plugs and ear muffs; reducing noise up to 30-40 dB). Additionally, non-standard MR acquisition schemes such as “silent” or “clustered” imaging can provide a solution (Eden et al., 1999; Gaab et al., 2007). In “silent” (or “sparse”) acquisition schemes, long TRs are used (12s – 20s). Because of the slow dynamics of BOLD responses, measured responses are not contaminated by scanner noise (Hall et al., 1999; Formisano et al., 2003). Silent acquisition schemes involve long measurement sessions, as they are statistically inefficient. Alternatively, in clustered imaging the actual image acquisition (and thus the scanner noise) is limited to a smaller initial fraction of TR (acquisition time or TA), leaving a silent gap between two subsequent volume acquisitions that can be used for stimulus presentation against a silent background. While clustered imaging does not eliminate the potential confounds derived from scanner noise-induced activation of the auditory cortex, it regains some of the statistical power compared to a silent acquisition scheme and therefore represent a good compromise. Clustered acquisition schemes are used in all experiments of the current thesis.



**Figure 5. Encoding model: describing voxels in terms of stimulus features**

Four stages are characteristic of encoding approaches. First, a model representing sounds in terms of their features is selected. Second, responses of the brain to a set of 'training' and 'testing' stimuli are measured. Third, machine-learning techniques are used to estimate the relationship between the features and the fMRI responses acquired for the training data. Finally, the learned functional relationship can be used to predict the pattern of activity of an entirely new stimulus.



## Encoding models

To date, the sensory sound representation in the human brain and the computational mechanisms that underlie the transition from an acoustically faithful sound image into a more abstract, behaviourally relevant and categorical sound representation remain largely elusive. In the current thesis, these topics are investigated by adopting methods of ‘receptive field’ mapping gleaned from electrophysiological recordings in non-human primates. Neuronal receptive fields - or in the case of fMRI, neuronal population profiles - can be estimated with synthetic or natural stimuli. Synthetic stimuli, while allowing for a careful control over acoustic features, suffer two disadvantages. First, they lack the complexity and statistical properties of natural sounds. As neurons have nonlinear receptive fields, these receptive fields are not estimated well with simple synthetic sounds (Theunissen et al., 2000; Bitterman et al., 2008). Second, synthetic stimuli lack ecological validity and familiarity, and consequently these sounds may not activate optimally the cortical auditory regions that are crucial to the transformation from a sensory into a more abstract sound representation. As a result, natural sounds are optimal for studying the research questions of the current thesis.

In this thesis, an ‘encoding’ model is used to describe brain locations by their tuning to acoustic features based on responses to natural sounds (Kay et al., 2008; Naselaris et al., 2010). Encoding models provide an explicit, quantitative description of how features of natural stimuli are represented in the activity of single voxels. Four stages are characteristic of encoding approaches (see Figure 5). First, a model that represents the stimuli in terms their features is selected. In the studies of the present thesis, sounds are described by their spectral energy in frequency bins, based on a model of early auditory processing (described in Chi et al. 2005; 40 frequency bins in chapters 2 and 3, 128 bins in chapter 4, and 64 bins in chapter 5). Second, responses of the brain to a set of ‘training’ and ‘testing’ stimuli are measured. Third, machine-learning techniques are used to find a functional relationship between the features and the fMRI responses acquired for the training data set (i.e. “encoding”). Finally, the learned functional relationship can be used to predict new patterns of stimulus-induced brain activity, evoked by independently measured stimuli (i.e. not used for defining the model; Kay et al., 2008; Naselaris et al., 2009). If prediction accuracy is significantly above chance, it is concluded that the estimated feature model – consisting of a neuronal population profile for each voxel - describes a relevant part of brain processing. Note that in this manner different models can directly be compared to each other. Subsequently, the neuronal population profiles can be interpreted, for example by creating topographic

maps of encoded features (i.e. tonotopic maps extracted in each of the chapters, or tuning width maps extracted in chapters 2 and 3) or by clustering the profiles to reveal characteristic types of feature tuning (see chapters 4 and 5).

## **Research aims and outline of the thesis**

Two general research questions are addressed in this thesis. First, how is the spectral content of natural sounds represented in our brain? Second, what mechanisms underlie the transition of a tonotopic sound image into a more abstract, behaviourally relevant spectral representation?

Chapter two investigates spectral tuning in the human auditory cortex by mathematical modelling of fMRI responses to natural sounds. The functional topography of the human superior temporal plane is defined by topographic maps of frequency preference (tonotopy) and selectivity (tuning width). Furthermore, the study in chapter two reveals that there is an intrinsic bias to low sound frequencies in categorical regions sensitive to human speech and voices. The study described in chapter three exploits the advantages of ultra-high field imaging to map, for the first time in humans, the detailed functional organization of the inferior colliculus. Results reveal a topographic organization of both frequency (tonotopy) and tuning width. Furthermore, this study provides a direct comparison between human subcortical and cortical spectral tuning. The fourth chapter uses an innovative data-driven approach to characterize cortical spectral tuning beyond its main tonotopic peak. The observed neuronal population sensitivity to relevant combinations of frequency bands is suggested to serve in detecting complex sound features, aid in segregating auditory scenes, and explain prominent perceptual phenomena such as octave invariance. The fifth chapter tests the hypothesis formulated in chapter four, and shows that fine multi-peaked spectral tuning within human PAC underlies the perceptual phenomenon of octave invariance. Finally, chapter six summarizes the findings of these studies and proposes a model of human auditory cortex based on its spectral tuning properties.

## References

- Balaguer-Ballester E, Denham SL, Meddis R (2008) A cascade autocorrelation model of pitch perception. *J Acoust Soc Am* 124: 2186-2195.
- Bandyopadhyay S, Shamma SA, Kanold PO (2010) Dichotomy of functional organization in the mouse auditory cortex. *Nat Neurosci* 13: 361-368.
- Baumann S, Griffiths TD, Sun L, Petkov CI, Thiele A, Rees A (2011) Orthogonal representation of sound dimensions in the primate midbrain. *Nat Neurosci* 14: 423-425.
- Bear MF, Connors BW, Paradiso MA (2007) *Neuroscience: Exploring the Brain*. Baltimore: Lippincott Williams & Williams.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000). Voice-selective areas in human auditory cortex. *Nature* 403: 309-312.
- Bilecen D, Scheffler K, Schmid N, Tschopp K, Seelig J (1998) Tonotopic organization of the human auditory cortex as detected by BOLD-FMRI. *Hearing Research* 126: 19-27.
- Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I (2008) Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451: 197-202.
- Brodmann K (1909) *Vergleichende lokalisationslehre der Grosshirnrinde*. Leipzig: JA Barth.
- Campaigne R, Minckler J (1976) A note on the gross configurations of the human auditory cortex. *Brain Lang* 3: 318-323.
- Cavaco S, Lewicki MS (2007) Statistical modeling of intrinsic structures in impacts sounds. *J Acoust Soc Am* 121: 3558-3568.
- Celesia GG (1976) Organization of auditory cortical areas in man. *Brain* 99: 403-414.
- Chevillet M, Riesenhuber M, Rauschecker JP (2011) Functional correlates of the anterolateral processing hierarchy in human auditory cortex. *J Neurosci* 31: 9352-9345.
- Chi T, Ru P, Shamma SA (2005) Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am* 118: 887-906.
- Crandall IB, MacKenzie D (1922) Analysis of the energy distribution in speech. *The Physical Review* 19: 116-128.
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci* 31: 12067-14075.
- Di Salle F, Esposito F, Scarabino T, Formisano E, Marciano E, Saulino C, Cirillo S, Elefante R, Scheffler K, Seifritz E (2003) fMRI of the auditory system: understanding the neural basis of auditory gestalt. *Magn Reson Imaging* 21: 1213-1224.
- Eden GF, Joseph JE, Brown HE, Brown CP, Zeffiro TA (1999) Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. *Magn Reson Med* 41: 13-20.
- Engelien A, Yang Y, Engelien W, Zonana J, Stern E, Silbersweig DA (2002) Physiological mapping of human auditory cortex with silent event-related fMRI technique. *Neuroimage* 16: 944-953.
- Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science* 322: 970-973.
- Formisano E, Kim D, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40: 859-869.
- Gaab N, Gabrieli JD, Glover GH (2007) Assessing the influence of scanner background noise on auditory processing. I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. *Hum Brain Mapp* 28, 703-720.
- Galaburda AM, LeMay M, Kemper TL, Geschwind N (1978) Right-left asymmetries in the brain. *Science* 199: 852-856.
- Galaburda AM, Sanides F (1980) Cytoarchitectonic organization of the human auditory cortex. *The Journal of Comparative Neurology* 190: 597-610.
- Geschwind N, Levitsky W (1968) Human brain: left-right asymmetries in temporal speech region. *Science* 161: 186-187.

- Goense JB, Logothetis NK (2008) Neurophysiology of the BOLD fMRI signal in awake monkeys. *Curr Biol* 18:631-640.
- Golding NL, Oertel D (2012) Synaptic integration in dendrites: Exceptional need for speed. *J Physiol* 590: 5563-5569.
- Gygi B, Kidd GR, Watson CS (2007) Similarity and categorization of environmental sounds. *Perception & Psychophysics* 69: 839-855.
- Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441: 197-222.
- Hackett TA, Rinaldi Barkat T, O'Brien BMJ, Hensch TK, Polley DB (2011) Linking topography to tonotopy in the mouse auditory thalamocortical circuit. *J Neurosci* 31: 2983-2995.
- Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394: 475-495.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) "Sparse" temporal sampling in auditory fMRI. *Hum Brain Mapp* 7: 213-223.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Aschouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425-2430.
- Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. *Neuroimage* 50: 1202-1211.
- Joly O, Ramus F, Pressnitzer D, VanDuffel W, Orban GA (2011) Interhemispheric differences in auditory processing revealed by fMRI in awake rhesus monkeys. *Cerebral Cortex* 22: 838-853.
- Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. *Physiol Rev* 84: 541-577.
- Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci* 97: 11793-11799.
- Kajikawa Y, de La Mothe, Blumell S, Hackett TA (2005) A comparison of neuron response properties in areas A1 and CM of the marmoset monkey auditory cortex: tones and broadband noise. *J Neurophysiol* 93: 22-34.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17: 4302-4311.
- Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. *Nature* 452: 352-356.
- Killian-Huetten N, Valente G, Vroomen J, Formisano E (2011) Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J Neurosci* 31: 1715-1720.
- Kim JJ, Crespo-Facorro B, Andreasen NC, O'Leary DS, Zhang B, Harris G, Magnotta VA (2000) An MRI-based parcellation method for the temporal lobe. *Neuroimage* 11: 271-288.
- King AJ, Nelken I (2009) Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat Neurosci* 12: 698-701.
- Koopmans PJ, Barth M, Norris DG (2010) Layer-specific BOLD activation in human V1. *Human Brain Mapping* 31: 1297-1304.
- Kosaki H, Hashikawa T, He J, Jones EG (1997) Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. *J Comp Neurol* 386: 304-316.
- Kusmierek P, Rauschecker JP (2009) Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J Neurophysiol* 102: 1606-1622.
- Langers DR, van Dijk P (2012) Mapping the tonotopic organization of the human auditory cortex with minimally salient acoustic stimulation. *Cerebral Cortex*, doi: 10.1093/cercor/bhr282.
- Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J Neurosci* 30: 7604-7612.

- Lewis JW, Talkington WJ, Walker NA, Spirou GA, Jajosky A, Frum C, Brefczynski-Lewis JA (2009) Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *J Neurosci* 29: 2283-2296.
- Malmierca MS, Izquierdo MA, Cristaudo S, Hernandez O, Perez-Gonzalez D, Covey E, Oliver DL (2008) A discontinuous tonotopic organization in the inferior colliculus of the rat. *J Neurosci* 28: 4767-4776.
- Meddis R, Lopez-Poveda EA (2010) Auditory periphery: from pinna to auditory nerve. In: *Computational Models of the Auditory System* (Meddis R, Lopez-Poveda EA, Fay RR, Popper AN, eds), pp7-38. New York: Springer.
- Merzenich MM, Brugge JF (1973) Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res* 50: 275-296.
- Merzenich MM, Knight PL, Roth GL (1973) Representation of cochlea within primary auditory cortex in the cat. *Brain Res* 63: 343-346.
- Merzenich MM, Reid MD (1974) Representation of the cochlea within the inferior colliculus of the cat. *Brain Research* 77: 397-415.
- Morel A, Garraghty PE, Kaas JH (1993) Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J Comp Neurol* 15: 437-459.
- Morosan P, Rademacher J, Schleicher A, Amunts A, Schormann T, Zilles K (2001) Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13: 684-701.
- Morosan P, Schleicher A, Amunts K, Zilles K (2005) Multimodal architectonic mapping of human superior temporal gyrus. *Anat Embryol* 210: 401-406.
- Naselaris T, Kay KN, Nishimoto S, Gallant JL (2010) Encoding and decoding in fMRI. *Neuroimage* 56: 400-410.
- Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian Reconstruction of natural images from human brain activity. *Neuron* 63: 902-915.
- Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A* 89: 5951-5955.
- Oliver DL, Beckius GE, Bishop DC, Loftus WC, Bantra R (2003) Topography of interaural temporal disparity coding in projections of medial superior olive to inferior colliculus. *J Neurosci* 23: 7438-7449.
- Patterson RD, Robinson K, Holdsworth J, McKeown D, Zhang C, Allerhand M (1992) Complex sounds and auditory images. In: *Auditory Physiology* (Cazals Y, Horner K, Demany L, eds), pp429-443.
- Penhune VB, Zatorre RJ, MacDonald JD, Evans AC (1996) Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb Cortex* 6: 661-672.
- Petkov CI, Kayser C, Augath M, Logothetis NK (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol* 4: doi: 10.1371/journal.pbio.0040215.
- Philips DP, Irvine DRF (1981) Responses of single neurons in physiologically defined primary auditory cortex (A1) of the cat: Frequency tuning and responses to intensity. *J Neurophysiol* 45: 48-58.
- Polimeni JR, Fischl B, Greve DN, Wald LL (2010) Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *NeuroImage* 52: 1334-1346.
- Rademacher J, Caviness Jr VS, Steinmetz H, Galaburda AM (1993) Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping and neurobiology. *Cerebral Cortex* 3: 313-329.
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience* 12: 718-724.

- Rauschecker JP, Tian B (2004) Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91: 2578-2589.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111-114.
- Reale RA, Imig TJ (1980) Tonotopic organization in auditory cortex of the cat. *J Comp Neurol* 192: 265-291.
- Recanzone GH, Cohen YE (2010) Serial and parallel processing in the primate auditory cortex revisited. *Behavioural Brain Research* 206: 1-7.
- Remedios R, Logothetis NK, Kayser C (2009) Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. *Proc Natl Acad Sci U S A* 106: 18010-18015.
- Rivier F, Clarke S (1997) Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: Evidence for multiple auditory areas. *Neuroimage* 6: 288-304.
- Rodriguez FA, Read HL, Escabi MA (2010) Spectral and temporal modulation tradeoff in the inferior colliculus. *J Neurophysiol* 103: 887-903.
- Rothschild G, Nelken I, Mizrahi A (2010) Functional organization and population dynamics in the mouse primary auditory cortex. *Nat Neurosci* 13: 353-360.
- Schneggenburger R, Forsythe ID (2006) The calyx of Held. *Cell Tissue Res* 326: 311-337.
- Schönwiesner M, von Cramon DY, Rubsamen R (2002) Is it tonotopy after all? *Neuroimage* 17: 1144-1161.
- Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* 60: 1823-1840.
- Schreiner CE, Langner G (1997) Laminar fine structure of frequency organization in auditory midbrain. *Nature* 388: 383-386.
- Schreiner CE, Winer JA (2007) Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron* 56: 356-365.
- Singh NC, Theunissen F (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 114: 3394-3411.
- Staeren N, Renvall H, De Martino F, Goebel R, Formisano E (2009) Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology* 19: 498-502.
- Striem-Amit E, Hertz U, Amedi A (2011) Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLoS One* 6: DOI:10.1371/journal.pone.0017832.
- Sweet RA, Dorph-Petersen KA, Lewis DA (2005) Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *J Comp Neurol* 491: 270-289.
- Talavage TM, Ledden PJ, Benson RR, Rosen BR, Melcher JR (2000) Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hearing Research* 150: 225-244.
- Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, et al. (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol* 91: 1282-1296.
- Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315-2331.
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. *Science* 292: 290-293.
- Tollin DJ, Koka K, Tsai JJ (2008) Interaural level difference discrimination thresholds for single neurons in the lateral superior olive. *J Neurosci* 28: 4848-4860.
- Ungerleider L, Mishkin M (1982) Two cortical visual systems. In: *Analysis of Visual Behavior* (Engle DJ, Goodale MA, Mansfield RJ, eds.), pp549-586. Cambridge, MA: MIT Press.
- Vaughan JT, Garwood M, Collins CM, Liu W, DelaBarre L, Adriany G, Andersen P, Merkle H, Goebel R, Smith MB, Ugurbil K (2001) 7T vs. 4T: RF power, homogeneity, and signal-to-noise

- comparison in head images. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 46: 24-30.
- Wallace MN, Johnston PW, Palmer AR (2002) Histochemical identification of cortical areas in the auditory region of the human brain. *Exp Brain Res* 143: 499-508.
- Wenstrup JJ (2005) The tectothalamic system. In: *The Inferior Colliculus* (Winer JA and Schreiner CE, eds), pp200-230. New York: Springer.
- Winer JA and Schreiner CE (2005) *The Inferior Colliculus*. New York: Springer.
- Winer JA and Schreiner CE (2010) *The Auditory Cortex*. New York: Springer.
- Woods DL, Stecker GC, Rinne T, Herron TJ, Cate AD, Yund EW, Liao I, Kang X (2009) Functional maps of human auditory cortex: effects of acoustic features and attention. *PLoS One*. 4: doi:10.1371/journal.pone.0005183.
- Woods DL, Herron TJ, Cate AD, Yund EW, Stecker GC, Rinne T, Kang X (2010) Functional properties of human auditory cortical fields. *Frontiers in systems neuroscience* 4: doi: 10.3389/fnsys.2010.00155.
- Yacoub E, Schmuell A, Pfeuffer J, van de Moortele PF, Adriany G, Andersen P, Vaughan JT, Merkle H, Ugurbil K, Hu X (2001) Imaging brain function in humans at 7 Tesla. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 45: 588-594.
- Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: Music and speech. *Trends Cogn Sci* 6: 37-46.

## **Chapter 2**

### **Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning and relation to voice-sensitivity**

Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci* 32: 14205-14216.





## Abstract

Auditory cortical processing of complex meaningful sounds entails the transformation of sensory (tonotopic) representations of incoming acoustic waveforms into higher-level sound representations (e.g. their category). However, the precise neural mechanisms enabling such transformations remain largely unknown. In the present study, we use functional MRI and natural sounds stimulation to examine these two levels of sound representation (and their relation) in the human auditory cortex. In a first experiment, we derive cortical maps of frequency preference (tonotopy) and selectivity (tuning width) by mathematical modeling of fMRI responses to natural sounds. The tuning width maps highlight a region of narrow tuning that follows the main axis of Heschl's gyrus and is flanked by regions of broader tuning. The narrowly tuned portion on Heschl's gyrus contains two mirror-symmetric frequency gradients, presumably defining two distinct primary auditory areas. In addition, our analysis indicates that spectral preference and selectivity (and their topographical organization) extend well beyond the primary regions and also cover higher-order and category-selective auditory regions. In particular, regions with preferential responses to human voice and speech occupy the low frequency portions of the tonotopic map. We confirm this observation in a second experiment, where we find that speech/voice selective regions exhibit a response bias toward the low frequencies characteristic of human voice and speech, even when responding to simple tones. We propose that this frequency bias reflects the selective amplification of relevant and category-characteristic spectral bands, a useful processing step for transforming a sensory (tonotopic) sound image into higher level neural representations.

## Introduction

Natural sounds consist of various and complex temporal patterns of acoustic energy extending over a wide range of frequency bands. How does our brain deal with this variety and complexity? At the sensory periphery (cochlea) and in the sub-cortical auditory relays, the sound frequency bands are selectively processed in spatially segregated channels (King and Nelken, 2009). This frequency-selective processing is preserved in the cortical areas at the early stages of the auditory processing hierarchy. In these areas, neurons with similar frequency preference cluster together and form multiple cochleotopic or tonotopic maps (Merzenich & Brugge, 1973). These tonotopic maps are thought to encode sensory sound representations at different temporal and spectral resolutions, which may be used for efficient processing of auditory scenes (Elhilali & Shamma, 2008).

Beyond frequency selectivity, electrophysiology in animals and humans indicates that neural processing of natural sounds involves non-linear computational mechanisms (Bitterman et al., 2008; Pasley et al., 2012), such as tuning to the statistical regularities of conspecific vocalizations (Theunissen et al., 2000) and context dependence of receptive fields (David et al., 2009). Furthermore, converging evidence from research in human (Belin et al., 2000; Zatorre et al., 2002) and non-human (Petkov et al., 2008) primates suggest that the auditory cortex includes functionally specialized regions, where neuronal populations respond stronger to conspecific vocalizations than to various control sounds. These regional activations are thought to express “semantic” sound representations for which the relation to the acoustic make-up is partially lost (Belin et al., 2000; Perrodin et al., 2011). To date, it is unknown how such higher-level representations relate to low-level sensory representations.

The aim of this study is twofold. A first aim is to examine the topographic organization of spectral responses in human auditory cortex using fMRI and natural sounds stimulation. Compared to synthetic sounds (tones), ecologically valid sounds engage the auditory cortex in meaningful processing, thereby enabling optimal analysis of spectral tuning in both primary and non-primary auditory areas. We show that maps of frequency preference (tonotopy) and selectivity (tuning width) can be derived from mathematical modelling of fMRI responses to natural sounds and that this allows delineating cortical fields within and outside the primary auditory cortex.

Using natural sounds in the context of tonotopic mapping also enables a direct comparison between higher (e.g. category selectivity) and lower niveau (e.g. frequency tuning) response properties. Recent findings indicate that basic response properties of category-selective regions in the visual cortex (e.g. face-

[Hasson et al., 2003] or place- [Rajimehr et al., 2011] selective regions) are closely related to characteristic physical properties of their preferred stimuli. Thus, a second aim is to test whether a similar mechanism is in place in the auditory cortex. We show that speech/voice selective regions exhibit a response bias toward the low frequencies typical of human vocal sounds, even when responding to tones. We propose that this frequency bias reflects the selective amplification of informative spectral energy, a useful processing step for transforming a sensory (tonotopic) sound image into a higher-level neural representation.

## Methods

### Subjects

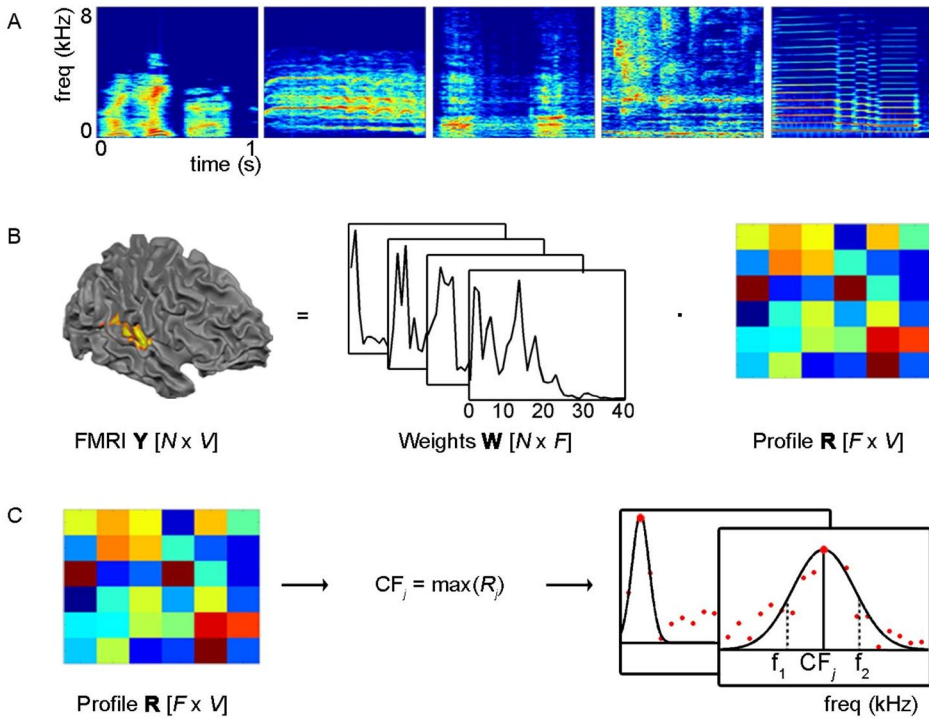
Five subjects (median age = 26, 3 males) participated in this study, which included two separate experiments ('natural sounds' and 'localizer', see below). The subjects had no history of hearing disorder or neurological disease, and gave informed consent before commencement of the measurements. The Ethical Committee of the Faculty of Psychology and Neuroscience at Maastricht University granted approval for the study.

### Stimuli

In the 'natural sounds' experiment, we used the recordings of various natural sounds as stimuli (Figure 1A). This set of sixty sounds included human vocal sounds (both speech and non-speech, e.g. baby cry, laughter, coughing), animal cries (e.g. monkey, lion, horse), musical instruments (e.g. piano, flute, drums), scenes from nature (e.g. rain, wind, thunder), and tool sounds (e.g. keys, scissors, vacuum cleaner). Sounds were sampled at 16 kHz and their duration was cut at 1000 ms.

In the 'localizer' experiment, we employed sounds grouped into eight conditions (3 tones and 5 semantic category conditions). For the tones conditions, amplitude modulated tones were created in Matlab (8 Hz, modulation depth of 1) with a carrier frequency of 0.45, 0.5, and 0.55 kHz for the low frequency condition; 1.35, 1.5, and 1.65 kHz for the middle frequency condition and 2.25, 2.5, and 2.75 kHz for the high frequency conditions. For the semantic category conditions, we collected sixty sounds (twelve sounds per condition; speech, voice, animals, tools, and nature sounds). Sounds were sampled at 16 kHz and their duration was cut at 800 ms.

In both the natural sounds and localizer experiment, sound onset and offset were ramped with a 10 ms linear slope, and their energy (RMS) levels were equalized. Inside the scanner - before starting the measurement - sounds were played to the subject while headphones and earplugs were in place. Intensity of the sounds was further adjusted in order to equalize their perceived loudness. As our goal was to present sounds as naturally as possible, no further manipulation of the stimuli was performed. In the natural sounds experiment, sounds were presented using an audio stimulation device with headphones developed by Resonance Technology Inc. ([www.mrvideo.com](http://www.mrvideo.com)). For the localizer, we presented the sounds using the S14 model fMRI-compatible earphones of Sensimetrics Corporation ([www.sens.com](http://www.sens.com)).



**Figure 1. Sound examples and computation of topographic maps.**

(A) Spectrotemporal representation of five exemplary sounds (i.e. speech, bird, dog, water, and flute) as output of the computational model mimicking early auditory processing. (B) Modelling of the measured brain responses  $\mathbf{Y}$  to the sounds' spectral components  $\mathbf{W}$  allows estimating – through regularized regression – the frequency response profiles  $\mathbf{R}$  of all cortical locations ( $N$  = number of sounds;  $F$  = number of frequencies;  $V$  = number of locations). (C) Characteristic frequency ( $CF$ ) of a location is obtained as the maximum of  $R_j$ . Tuning width ( $W$ ) is obtained as the relative width ( $CF_j/[f_2-f_1]$ ) of a Gaussian function fitted to estimated spectral profiles, while fixing the mean of the Gaussian to  $CF_j$ .

## Magnetic Resonance Imaging

Images were acquired on a 3T head only MR scanner (Siemens Allegra) at the Maastricht Brain Imaging Center (M-BIC). Anatomical  $T_1$ -weighted volumes covering the whole brain were obtained with an ADNI MPRAGE sequence (TR = 2250 ms; TE = 2.6 ms; matrix size = 256 x 256 x 192, voxel dimensions = 1 x 1 x 1 mm<sup>3</sup>). Functional  $T_2^*$ -weighted images were collected using a clustered volume EPI technique.

The natural sounds experiment was designed according to a fast event-related scheme. The acquisition parameters were: TR = 2600 ms; time of acquisition [TA] = 1200 ms; TE = 30 ms; number of slices = 13; matrix size = 128 x 128; voxel size = 2 x 2 x 2 mm<sup>3</sup>, silent gap = 1400 ms. Sounds were randomly spaced at a jittered interstimulus interval of 2, 3, or 4 TRs and presented – with additional random jitter - in the silent gap between acquisitions. Zero trials (trials where no sound was presented, 10% of the trials), and catch trials (trials in which the sound which was just heard was presented, 6% of the trials) were included. Subjects were instructed to perform a one-back task, and were required to respond with a button press when a sound was repeated. Per run, each of the sixty sounds was presented three times. Catch trials were excluded from the analysis. The full measurement session consisted of three runs (~25 min each).

The localizer experiment was designed according to a blocked scheme. The acquisition parameters were: TR = 3000 ms; time of acquisition [TA] = 1500 ms; TE = 30 ms; number of slices = 18; matrix size = 128 x 128; voxel size = 2 x 2 x 2 mm<sup>3</sup>, silent gap = 1500 ms). Sounds of the same condition were presented in blocks of six (one sound per TR, presented in the silent gap). Blocks of acoustic stimulation lasted 18 s and were separated by 12 s of silence. Per run, two blocks of each of the eight conditions were presented. The full measurement session consisted of six runs (~ 9 min each).

Functional and anatomical images were analyzed with BrainVoyager QX. Preprocessing consisted of slice scan-time correction (using sinc interpolation), temporal high-pass filtering to remove drifts of 3/11 or less cycles per time course (localizer and natural sounds experiment respectively), and 3-dimensional motion correction. Functional slices were co-registered to the anatomical data, and both datasets were normalized to Talairach space (Talairach & Tournoux, 1988). For the localizer only, temporal smoothing of 2 datapoints and moderate spatial smoothing with a 3 mm kernel were applied. Anatomical volumes were also used to derive gray matter segmentations indicating the border between white and gray matter. Using this border, inflated hemispheres of the individual subjects were obtained. Next, cortex-based alignment (CBA) was used to im-

prove alignment of the major sulci and gyri between subjects (Goebel et al., 2006). This alignment information was used for calculating and displaying group maps.

### Topographic maps from responses to natural sounds

We calculated frequency (CF) and tuning width (W) maps from the natural sounds experiment using customized Matlab code ([www.mathworks.com](http://www.mathworks.com)). We followed methodological procedures similar to the ones previously described for the analyses of visual responses to natural scenes (Kay et al., 2008b; Naselaris et al., 2009, 2010), and adapted them to the analysis of natural sounds.

As a first step, we analyzed the sounds used as stimuli in our natural sounds experiment with a biologically plausible computational model of auditory processing from the cochlea to the midbrain (Chi et al., 2005). Within the model, sounds are passed through a bank of filters with center frequency equally spaced on a logarithmic axis (128 filters, constant Q), to represent the cochlear filter output. Consequent operations mimic the computations performed by the auditory nerve and cochlear nucleus (Chi et al., 2005), resulting in the mathematical representation of the sounds in terms of an  $N \times F$  matrix  $\mathbf{W}$  of coefficients, where  $N$  = number of sounds, and  $F$  = the number resulting frequency bins ( $N = 60$  and  $F = 128$ ).  $F$  was resampled to 40 bins with the center frequencies spanning 5.2 octaves (bins uniformly distributed on a logarithmic frequency axis; center frequency ranges from 186 to 6817 Hz). The choice for forty frequency bins is a trade-off between frequency resolution and a manageable number of free parameters to be estimated for each of the voxels (see below).

To calculate the voxels' frequency profile and topographic maps, we used an fMRI 'encoding' procedure (Kay et al., 2008b; Naselaris et al., 2009, 2010). The following computations were repeated in each subject using three different splits of the data (run 1/2, run2/3, and run1/3). Resulting maps and profiles were obtained as the average across splits: First, matrix  $\mathbf{Y}$  [ $(N \times V)$ ,  $N$  = number of sounds;  $V$  = number of voxels] of the fMRI responses to the sounds was calculated using a voxel-by-voxel General Linear Model (GLM) analysis (Friston et al., 1995). For each voxel  $j$ , the response vector  $\mathbf{Y}_j$  [ $(N \times 1)$ ] was estimated in two steps. First, with deconvolution, the hemodynamic response function (HRF) common to all stimuli (i.e. all stimuli were treated as a single condition) was estimated. Then, keeping the HRF shape fixed, we computed the response of each voxel to the individual sounds (as a beta weight) using one predictor per sound (Kay et al., 2008a). Voxels that showed a significant response to the sounds were chosen for further analysis ( $p < 0.05$ , uncorrected in order not to be too stringent at this stage of the process). Second, a matrix  $\mathbf{R}$  [ $F \times V$ ] of the

frequency response profiles of the voxels was calculated using the fMRI response matrix  $\mathbf{Y}$  [ $N \times V$ ] and the frequency representation of the sounds  $\mathbf{W}$  [ $N \times F$ ] (Figure 1B). For each voxel  $j$ , its frequency profile  $R_j[(F \times 1)]$  was obtained as the relevance vector machine (RVM) solution to the linear problem:

$$\mathbf{Y}_j = \mathbf{W} \cdot \mathbf{R}_j$$

where each element  $i$  of the vector  $\mathbf{R}_j$  describes the contribution of the frequency bin  $i$  to the overall response of voxel  $j$ . We computed maps of tonotopy by considering the characteristic frequency (CF) of a voxel as the maximum of the coefficients in  $\mathbf{R}_j$ . In order to compute tuning width we fitted a Gaussian function to the estimated response profile  $\mathbf{R}_j$  while fixing the mean of the Gaussian at the voxels' characteristic frequency ( $CF_j$ ). The tuning width of each significant voxel was calculated as  $W$ :

$$W = CF / (f_2 - f_1)$$

where  $(f_2 - f_1)$  corresponds to the width in Hz at FWHM of the Gaussian fit (Figure 1C). Higher  $W$ -values correspond to more narrowly tuned voxels, lower  $W$ -values correspond to more broadly tuned voxels.

Note that, per data split and voxel, we solved a linear system with 60 observations (brain responses to sounds) and 40 variables (frequency bins). This linear system can be solved by minimum least square fitting (e.g. General Linear Model), which would allow simple statistical assessment of the weights with parametric statistics. However, we used a regularized regression approach (RVM) to robustly deal with the collinearity of the design matrix ( $\mathbf{W}$ ). Because we use linear RVM and the number of variables is smaller than the number of observations, overfitting is not a concern.

To assess the statistical significance of the maps of tonotopy, we used permutation statistics as RVM regression does not allow for the use of simple parametric testing. In particular, for each voxel we estimated the empirical null-distribution of  $CF_j$  by randomly permuting (200 permutations per data split) the voxels' responses to all sounds and repeating the fitting procedure. That is, per permutation,  $N$  in matrix  $\mathbf{Y}$  was shuffled. Per voxel, we determined the significance of  $CF_j$  by counting the number of occurrences in which the permuted fits resulted in a higher value than the non-permuted estimates. Voxels for which this count resulted in less than 5 percent of the total number of permutations were considered significant.

Note that for the computation of tuning width maps, not only the high coefficients in  $\mathbf{R}_j$  are of interest (as was the case for assigning significance to the estimation of  $CF_j$ ). Instead, the full profile is of importance and thus we statistically assessed the estimates of voxels' tuning width as follows. For each voxel, we assigned significance by comparing the goodness-of-fit of the fitted Gaussian function to  $\mathbf{R}_j$  against a null-distribution obtained by repeating the Gaussian



fitting procedure for each of the 200 permutations. Voxels with a goodness-of-fit within or higher than the best 5 percent of the permutations were considered significant.

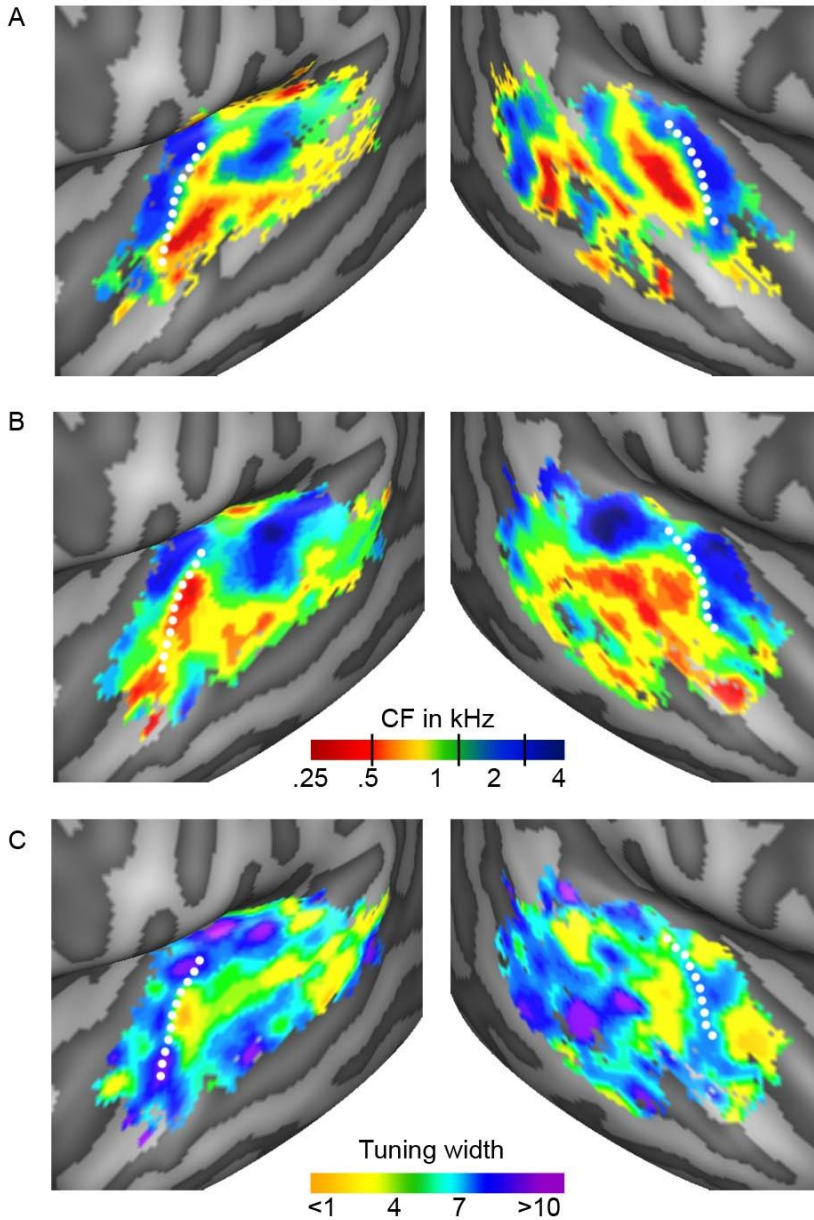
These analyses resulted in a map of tonotopy and tuning width for each of the three data splits, which were averaged to create one map of tonotopy and tuning width per subject. Group maps were computed by first transforming the individual maps to CBA-space. Voxels that had significant feature estimation in at least 3 out of the 5 subjects were included in the final group maps, which were obtained by averaging the individual subject maps in the CBA-space.

Tonotopic cortical maps were obtained by logarithmic mapping of best-frequency values to a red-yellow-green-blue colour scale, and tuning width maps were obtained by a linear mapping of tuning width values to a yellow-green-blue-purple colour scale (see Figure 2 and Figure 3). Although our stimuli included frequencies up to 8 kHz, only few voxels were assigned with such high CF. We may fail to observe selectivity to such high frequencies because natural sounds are mostly dominated by lower frequencies. Alternatively, frequencies above 4 kHz may evoke only weak fMRI responses. Further research is needed to specifically examine the fMRI responses to these high frequencies.

A correlation analysis between maps of tuning width and the voxels' overall response strength showed that these maps were unrelated to each other. However, as expected based on previous studies in animals (Cheung et al., 2001; Imaizumi et al., 2004), tonotopy and tuning width showed a significant positive correlation in each subject. Corrected maps of tuning width were computed as the residuals from fitting CF dependence of  $W$  with a smoothing spline. Corrected maps of tuning width displayed the same large-scale pattern as uncorrected maps. Consequently, uncorrected maps were used in the remainder of the analysis.

### Computation of unbiased topographic maps

In order to ensure that estimated CF and  $W$  values were not confounded by the frequency content of sound categories, we re-computed maps of tonotopy and tuning width on a subset of sounds with controlled frequency content across categories. Specifically, the voxels' response profile  $\mathbf{R}$  was calculated using the reduced matrices  $\mathbf{W}' [N' \times F]$ , and  $\mathbf{Y}' [(N' \times V)]$ , obtained from the full matrices  $\mathbf{W} [N \times F]$  and  $\mathbf{Y} [(N \times V)]$  by removing seven sounds (two low speech sounds, two low nature sounds, 2 high tool sounds, and one high animal sound;  $N' = 53$ ). All other steps remained identical to those described above. Exclusion of seven sounds from the analysis eliminated significant differences in center of gravity (CoG) across sound categories (assessed with independent samples t-tests).



**Figure 2. Group topographic maps**

Individual maps of tonotopy from the localizer (A) and the natural sounds (B) experiment. (C) Tuning width maps (natural sounds experiment), derived as  $CF \text{ (Hz)} / [\text{width of main peak (Hz)}]$ . Consequently, high values (in blue/purple) correspond to narrow tuning, and low values (in yellow/orange) correspond to broad tuning. The group topographic maps represent the mean across individuals and are shown for voxels included in  $\geq 3$  individual maps. White dotted lines indicate the location of HG.

### Topographic maps from responses to tones

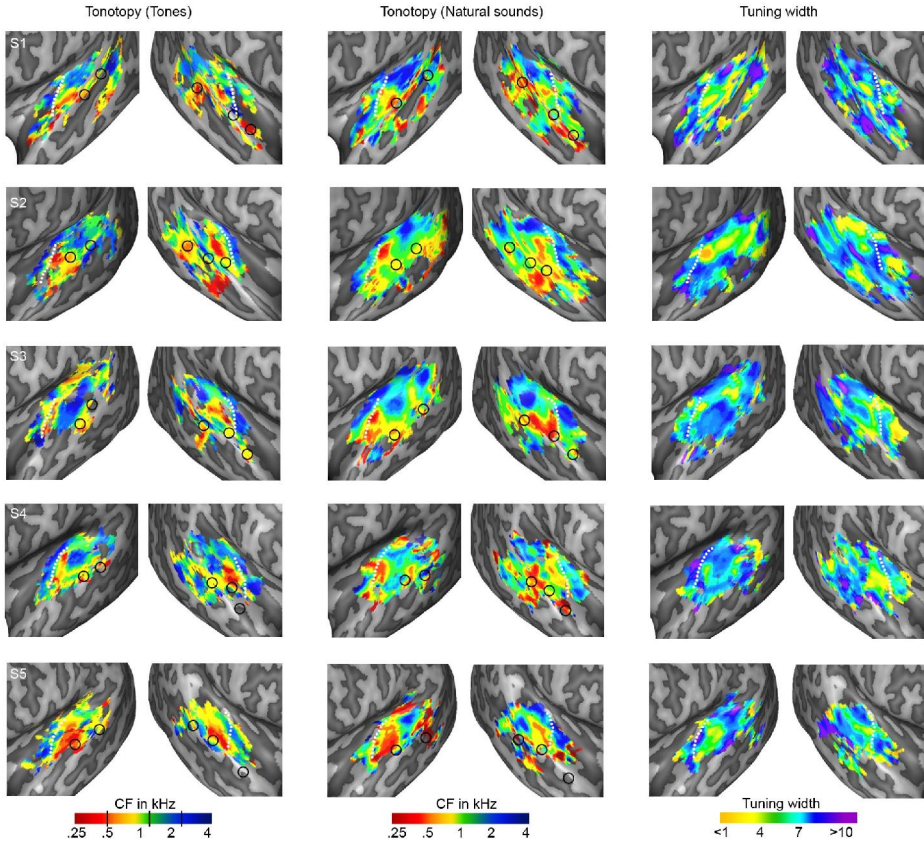
For comparison and validation, we also computed tonotopy ('best frequency maps') as elicited by the amplitude-modulated tones (Formisano et al., 2003). A single subject General Linear Model (GLM) analysis using a standard hemodynamic response model (Friston et al., 1995) was used to compute the responses to the three centre frequencies (.5; 1.5; 2.5 kHz) in all six runs separately. Voxels that showed a significant response to the sounds were selected ( $Q[FDR] < 0.05$ ), and response to the three tones was z-normalized across these voxels. For each voxel, its best frequency was determined in 6-fold cross-validation (one run was left out in each fold). If the estimated best frequency had a majority across folds (3 or more occurrences), the voxel was colour-coded accordingly. Colour-coding of best-frequency values was done using a red-yellow-green-blue colour scale.

We quantified the consistency between tonotopic maps extracted from natural sounds and tones with two analyses. First, for each subject we tested whether the global maps were significantly correlated to each other. We correlated the natural sound tonotopic map to the tones tonotopic map, and compared this value to the null-distribution obtained by correlating the natural sounds map to the permuted tones maps ( $N = 1000$ ). Significance was assigned at single-subject level, by counting the number of occurrences that the correlation to a permuted map was higher than the correlation to the unpermuted data. In order to evaluate this across-experiment correlation, we computed the correlation between maps resulting from the three different splits of the data in the natural sounds experiment (natural sounds-natural sounds correlation; run 1/2, run2/3, and run1/3) and between two maps resulting from half of the data in the tones experiment (tones-tones correlation; run 1/3/5, and run 2/4/6).

Second, we evaluated relative variations of consistency throughout the superior temporal cortex by computing difference maps comparing group tonotopy as extracted by tones and by natural sounds. In order to account for differences in estimated frequencies in the tones and natural sounds (ns) maps (three frequency bins vs. range of 5.2 octaves), we first normalized each map to values between 0 and 1. Then, we computed for each voxel  $j$  the normalized difference  $diff_j$  as:

$$diff_j = \frac{|tones_j - ns_j|}{tones_j + ns_j}$$

Note that a maximum difference in frequency between the two maps results in  $diff_j = 1$ . We compared median normalized difference across three anatomically



**Figure 3. Individual topographic maps.**

Individual maps of tonotopy (localizer and natural sounds experiment) and tuning width (natural sounds experiment) are shown. Note that the tonotopic maps are highly reproducible across experiments. White lines indicate the location of HG in individual subjects. Black circles outline the peaks of voice sensitive regions, defined at individual subject level based on responses in the natural sounds experiment (left column) and localizer experiment (middle column; see Figure 6).

defined regions: Heschl's gyrus (HG), planum temporale (PT), and superior temporal gyrus and sulcus (STG). Within each of these regions, we determined whether the observed difference was significantly lower than chance by comparing the median normalized difference to an empirically obtained null-distribution (null-distribution created by permuting the tones tonotopy map within each region, and computing median normalized difference to the natural sounds tonotopy map [ $N = 1000$ ]).

### Analysis of tonotopic gradients and field sign mapping

We determined the direction of the tonotopic gradient and the borders of the primary core regions by analyzing the obtained tonotopic group maps with a field sign mapping procedure (Formisano et al., 2003), modified to use the additional information from tuning width maps. First, we calculated frequency gradient maps by computing locally the direction of increasing frequency at each vertex of a flattened cortical mesh. Second, we used the tuning width maps to define a “reference” axis for the core region. Specifically, we defined the reference axis as the main orientation of the narrowly tuned region surrounding HG (see below). Next, we color-coded the gradient direction obtained at each vertex using a binary code: blue if the direction of the vertex was included within a  $\pm 90^\circ$  interval centered around the reference axis; green if the direction of the vertex corresponded to a mirror representation of the same interval.

To avoid potential biases in estimating the tonotopic gradient direction with a manual selection of the reference axis, this latter was defined analytically from the tuning width maps. In each hemisphere separately, we projected the tuning width map on the flattened cortical mesh including HG and surrounding gyri/sulci (FTS, HS, parts of lateral sulcus and STG) and constructed 3-dimensional (3D) data set (x-coordinate, y-coordinate, tuning width map value). We estimated our reference axis from the first eigenvector resulting from the singular value decomposition (SVD) of this 3-D data set. Furthermore, using an identical analysis, we calculated the principal “anatomical” axis of HG considering the anatomical gyral curvature as map value in the 3D data set. In the results, the relative angles between the “reference” tuning width axis and the anatomical axis of HG are reported.

### Definition of speech/voice sensitive regions

Individual regions sensitive for speech and voice sounds were defined using either localizer data or the data of the natural sounds experiment, by contrasting responses to speech and voice sounds with responses to all other natural sounds (animal cries, tool sounds, and nature sounds). When the natural sounds experiment was used to define speech/voice regions, further analyses were performed on the 300 most speech/voice selective voxels (average threshold ( $t$ ) = 2.8;  $p < 0.05$  in each subject). Alternatively, when the localizer was used to define speech/voice selective regions, a threshold at individual subject level of  $Q[FDR] < 0.001$  was used.

Five regions of speech/voice sensitivity were defined at individual subject level. For each subject, we defined a cluster on middle STG/STS (at the lateral

extremity of HS) bilaterally. Another bilateral cluster could be observed on posterior STG/STS. Finally, we identified a speech/voice cluster on the anterior STG/STS of the right hemisphere (at the lateral adjacency of HG/FTS). Although we also observed speech/voice selectivity on anterior STG/STS in the left hemisphere of some subjects, this region was not present in the majority of subjects and was therefore not included in the analysis. In order to visualize the location of speech/voice sensitive regions, individual regions were transformed to CBA-space and used to compute probabilistic maps.

### **Frequency bias and spectral tuning in speech/voice sensitive regions**

After defining speech/voice regions on the natural sounds experiment, we evaluated the frequency bias based on responses of the localizer. For each subject, responses to the eight sound conditions in the localizer (three tone and five natural sound conditions) were computed using a General Linear Model (GLM) analysis with a standard hemodynamic response model (Friston et al., 1995). Voxels that showed a significant response to the tones were selected ( $Q[FDR] < 0.05$ ), and the response to each condition was normalized across these voxels. Note that this normalization ensures that the average response to any condition across the auditory responsive cortex is equal to zero, excluding the possibility that deviations from zero are caused by an overall cortical response bias. Next, for each subject the average response to each condition in each speech/voice region was computed (i.e. middle and posterior STG/STS bilaterally, and anterior right STG/STS). For each region, the low frequency bias was quantified as the paired t-statistic comparing the response to low tones vs. the average of middle and high tones across subjects.

We performed our main analysis exploring the frequency bias in higher-level regions, in cortical areas defined by the contrast 'speech/voice vs. other sounds'. This choice was based on the observations that (1) in natural settings processing of speech necessarily involved processing of voice, and (2) overlapping regions in the superior temporal cortex are involved in both speech and voice processing (Formisano et al., 2008; Petkov et al., 2009). However, to explore the consistency of the low frequency bias for speech-sensitive and voice-sensitive regions separately, we repeated this analysis while defining regions of interest on the natural sounds experiment as (1) speech-sensitive, by contrasting responses to speech sounds vs. other sounds (i.e. animal, tool, nature sounds), and (2) voice-sensitive, by contrasting responses to voice sounds vs. other sounds (i.e. animal, tool, nature sounds). Here, the speech- and voice-sensitive region were analyzed as a whole (i.e. no division into separate clusters), by quantifying the low frequency bias as the paired t-statistic comparing

the response to low tones vs. the average of middle and high tones across subjects.

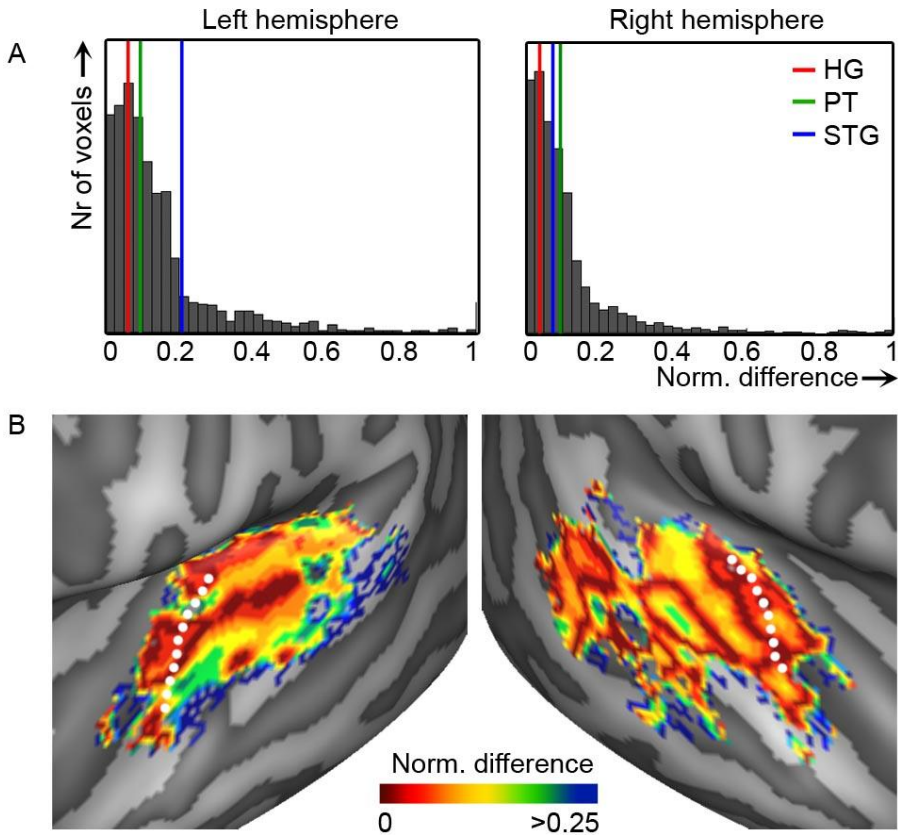
Next we defined speech/voice regions using the localizer and evaluated the full spectral profile based on the natural sounds experiment. Here, the spectral profile was either computed on all sounds or on the set of sound with equal CoG across categories. In order to quantify the low frequency bias, we divided the full spectral profile into eight bins equally spaced on a logarithmic frequency scale. Within each speech/voice region, we quantified the low-frequency bias by computing a paired t-statistic comparing across subjects the averaged responses to the three lowest frequency bins (maximum center frequency = 567 Hz) to the averaged responses of the remaining five frequency bins.

## Results

### General description of tonotopy and tuning width maps

As expected - both in the natural sounds and in the localizer experiment - sounds evoked significant activation in a large expanse of the superior temporal cortex. The activated region included early auditory areas along the Heschl's gyrus (HG) and surrounding regions on the planum polare (PP), planum temporale (PT), superior temporal gyrus (STG), and parts of the superior temporal sulcus (STS;  $Q[FDR] < 0.05$ ).

Figure 2 shows group maps of tonotopy and tuning width as extracted from the natural sounds experiment (Figure 2B-C) and from the localizer experiment (tonotopy only, Figure 2A; see Figure 3 for individual subject results and Figure 4 for a detailed comparison of tonotopic maps from natural sounds and tones). In the cortical regions surrounding HG (dotted white line), there was a clear spatial gradient of frequency preference. Regions preferring low frequencies (red-yellow color in Figure 2A-B) occupied the central part of HG and extended laterally and anteriorly. These low frequencies regions were surrounded medially and posteriorly by regions preferring high frequencies (green-blue color in Figure 2A-B). The medial high frequency cluster covered part of HG and the first transverse sulcus (FTS), and the posterior high frequency cluster extended from Heschl's sulcus (HS) throughout the PT. Anterior to the Heschl's region, on PP/anterior STG, an additional low frequency cluster was observed. Posterior to the Heschl's region, we observed clusters of low frequency preference (middle and posterior STG/STS) adjacent to regions preferring higher frequencies (on PT and middle STG; Figure 2A-B and Figure 5B).



**Figure 4. Consistency between tones and natural sounds tonotopy maps**

(A) In each hemisphere, normalized difference across tonotopic maps ranged mostly between 0 and 0.25. Consistency was largest on HG (red line), followed by PT (green line) and STG/STS (blue line). Note that in the left STG/STS, relatively large differences across experiments were observed. (B) These maps display the difference between normalized group tonotopy maps as extracted by tones and natural sounds across the superior temporal plane. Minimum and maximum difference between maps result in values of 0 (in red) and 1 (in blue) respectively.

Both at group level (Figure 2C) and in single subject maps (Figure 3, right column), tuning width maps presented a narrowly tuned region in the vicinity of HG (blue-purple color in Figure 2C). Surrounding this narrow region, areas with broader tuning were present (yellow-green in Figure 2C). These broadly tuned areas were located medially and laterally to HG in both hemispheres, and at the medial border of HG and HS in the right hemisphere. On PT and along STG/STS, distinct clusters of both narrow and broad tuning could be discerned.



In particular, two clusters with narrow tuning occupied the middle portion of STG/STS in symmetrical positions of left and right hemisphere.

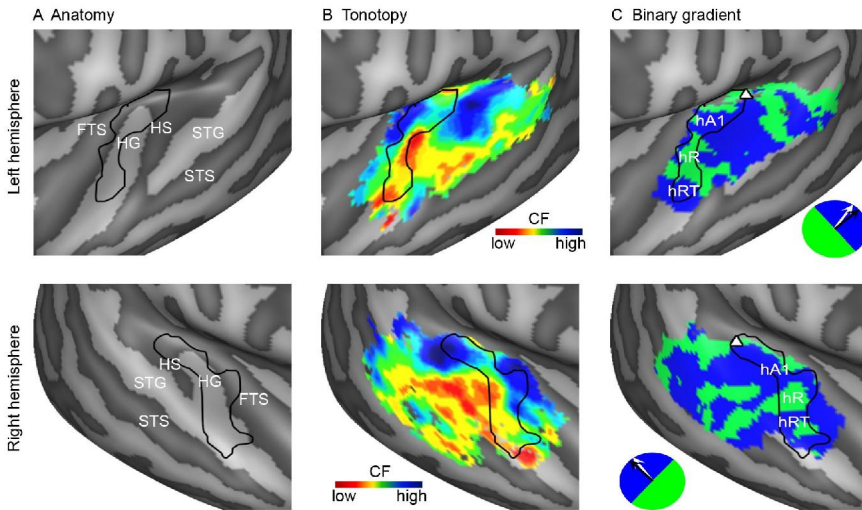
### Comparison of tonotopic maps obtained with natural sounds and tones

We compared the large-scale tonotopic pattern obtained from responses to natural sounds to that obtained with tones, both at group- (Figure 2A and Figure 2B) and at single-subject level (Figure 3, columns 1 and 2). For each subject, the spatial correlation between the two tonotopic maps was significantly higher than chance (mean[SD] correlation between maps = 0.18[0.05]; mean[SD] correlation permuted maps =  $1.29 \cdot 10^{-4}$ [ $8.06 \cdot 10^{-4}$ ];  $p < 0.001$  in each subject), suggest a relative similarity between maps. However, within-experiment correlation was noticeably higher than across-experiment correlation (natural sounds-natural sounds correlation mean[SD] = 0.74[0.02]; tones-tones correlation mean[SD] = 0.87[0.01]), reflecting the presence of differences between the tonotopic maps based on the two experiments.

Because the spatial correlation only provides a global measure of similarity, we computed difference maps that illustrate spatial (vowel-wise) variations in consistency throughout the superior temporal cortex (Figure 4). Values in this map range mostly between 0 and 0.25 (86% and 92% of voxels in left and right hemisphere respectively had a value below 0.25; Figure 4A) confirming a relative consistency. In both left and right hemisphere, highest consistency (red line in Figure 4A; red color in Figure 4B) was observed in regions on HG (median normalized difference was 0.06 and 0.03 for left and right hemisphere respectively;  $p < 0.001$ ). In the left hemisphere, consistency between tonotopic maps obtained with natural sounds and tones was high on PT (green line in Figure 4A; median normalized difference = 0.09,  $p < 0.05$ ) and decreased at the map extremities (i.e. STG/STS; blue line in Figure 4A; median normalized difference at chance level). Specifically, while regions of both low and high frequency preference were observed on STG/STS as estimated with natural sounds, this region was mainly tuned to low frequencies as estimated with tones. In the right hemisphere, this pattern reversed. Here, STG/STS displayed higher similarity across experiments than PT (median normalized difference was 0.07 [ $p < 0.05$ ] and 0.09 [ $p < 0.10$ ] respectively).

### Analysis of tonotopic gradients and field sign mapping

Figure 5 shows the results of a detailed analysis of the tonotopic gradients as obtained in the natural sounds experiment. Based on previous results in the monkey (Rauschecker et al., 1995; Rauschecker & Tian, 2004; Kajikawa et al.,



**Figure 5. Tonotopic field sign maps**

(A) For left and right hemisphere (top and bottom row respectively), the anatomy is shown. The narrowly tuned regions (i.e. high values in the maps of tuning width) in left and right hemisphere are outlined in black, showing that the main axis of these narrow regions is oriented along the main axis of HG. (B) Group tonotopic maps as extracted from natural sounds (also displayed in Figure 2B). The black outline shows the narrowly tuned region. (C) Field sign maps based on the tonotopic maps shown in (B). Blue and green colors indicate low-to-high and high-to-low tonotopic gradients with respect to the reference axis (i.e. the main axis of the narrowly tuned region, outlined in black) respectively. The reversal of the field sign defines the border between tonotopic maps (e.g. border between hA1 and hR). The reference axis is displayed by the black arrow in the circular blue-green color legend, and the main anatomical axis of HG is shown in white. The triangle indicates an additional frequency reversal at the medial convergence of HG and HS.

2005; Kusmirek & Rauschecker, 2009), we outlined the narrowly tuned region on HG as the primary auditory core (black outline in Figure 5A-C). The quantitative comparison of the principal axis of this region with the principal axis of HG (see Methods) indicated their almost parallel extension with a relative angle of  $13.6^\circ$  and  $8.4^\circ$  in left and right hemisphere respectively.

This primary core region included clusters with both low and high frequency preference (Figure 5B). Our field sign mapping analysis (Figure 5C) revealed - consistently in the left and right hemisphere - a “blue” region in the posterior medial part of HG (hA1) where the direction of the local “low-to-high” frequency gradient follows the direction of the reference axis (black arrow in the insert). Anterior to this blue region, a “green” region (hR) presented a reversed local gradient direction, confirming the existence of two bordering mirror-symmetric tonotopic maps (Formisano et al., 2003). In planum polare - anterior to this main tonotopic gradient - we observed another reversal of frequency preference. This frequency gradient, indicated in blue in Figure 5C, was also located in an area

of narrow tuning and possibly corresponds to the human homologue of area RT in the macaque (Kosaki et al., 1997; Hackett et al., 1998).

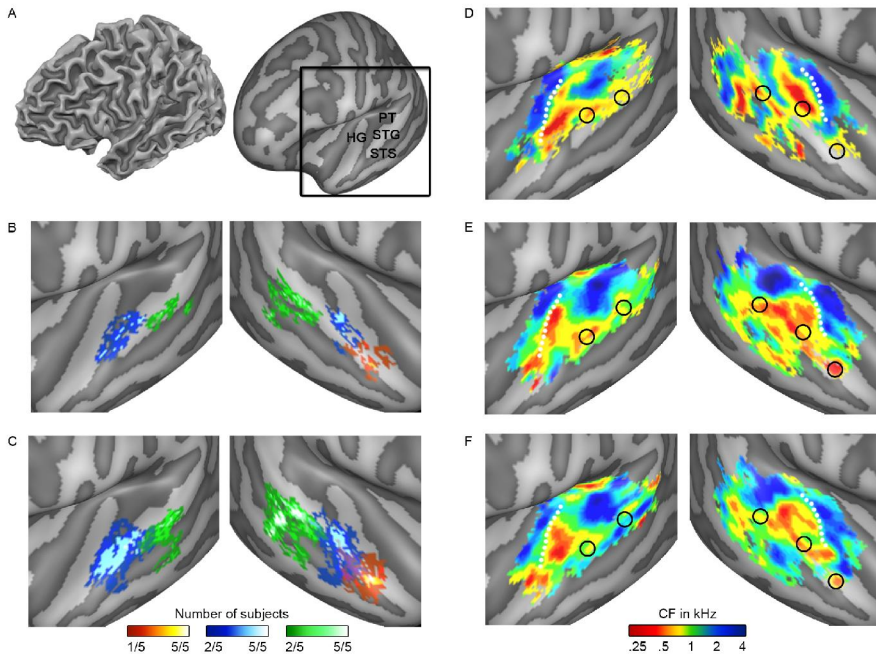
The field sign mapping analysis outside this core region suggested a complex arrangement of non-primary auditory areas. Towards the medial and posterior extremity of the narrowly tuned core region, extending into more broadly tuned parts of cortex, a green region was present bilaterally. This indicates the presence of another frequency reversal at the medial extremity of HG and HS (shown in green in Figure 5C and indicated by a triangle). At the posterior-lateral border of the narrow region, the frequency gradients tended to parallel those of the adjacent core areas and continued into broadly tuned regions of HS and further into PT and STG/STS. Especially in the right hemisphere, the direction of these gradients created a striped-like pattern in the field sign maps throughout the temporal cortex.

### **Frequency preference in speech/voice sensitive regions**

As the topographic maps extended into regions previously characterized as 'speech/voice sensitive' (Belin et al., 2000; Petkov et al., 2008), we examined in detail the relation between frequency preference and category sensitivity. We analyzed the data from our two experiments in two directions. First, we defined the speech/voice sensitive regions on the natural sounds experiment and analyzed the responses measured in the localizer. Second, we defined speech/voice sensitive regions on the localizer and analyzed the responses measured in the natural sounds experiment.

Regions that responded preferably to voice and speech sounds in the localizer and the natural sounds experiment showed a high degree of consistency with each other and with previous reports (Belin et al., 2000; Petkov et al., 2008). In both cases, we identified a bilateral speech/voice-sensitive cluster on middle STG/STS and posterior STG/STS, and an additional speech/voice-sensitive cluster on the anterior STG/STS in the right hemisphere (see Figure 6B-C, and Table 1 for average Talairach coordinates).

Speech/voice regions, as defined by the contrast speech/voice vs. other in the natural sounds experiment, presented a typical response profile in the localizer experiment (Figure 7). As expected, responses were stronger to speech and voice sounds than to other sound categories. Most interestingly, all regions



**Figure 6. Overlay between tonotopy and speech/voice regions**

(A) The black square outlines the cortical region shown in the remainder of this figure. The region includes Heschl's gyrus (HG), superior temporal gyrus (STG), superior temporal sulcus (STS), and planum temporale (PT). (B-C) Speech/voice regions as defined by the natural sounds experiment and by the localizer respectively. In both the natural sounds experiment and the localizer, we defined a middle (blue) and posterior (green) speech/voice cluster bilaterally. In the right hemisphere, we additionally identified an anterior cluster (red). For Talairach coordinates of speech/voice regions, see Table 1. (D) Tonotopy group maps as extracted from the localizer and (E-F) the natural sounds experiment (based on all sounds, and based on sounds with matched spectral content across categories respectively). In D-F, peaks of speech/voice sensitive regions are outlined in black on the tonotopy maps. See Figure 3 for all individual overlays and Table 1 and Figure 7 for the statistical assessment of the low-frequency bias in speech/voice regions.

occupied low frequency areas ( $<1$  kHz) in the tonotopic maps obtained with the localizer (Figure 6D) and responded more strongly to low frequency tones than to middle and high frequency tones in the localizer experiment (black dots in Figure 7). This pattern was present in each subject and in each region (for individual subject results, see column 1 in Figure 3). Statistical comparison of the responses to low tones vs. the averaged response to middle and high tones indicated a significant low frequency bias in all speech/voice sensitive clusters, except for the posterior cluster in the left hemisphere (see Table 1). When defining speech-specific and voice-specific regions using separate contrasts in the natural sounds experiment (i.e. speech vs. other and voices vs. other, respec-

tively), results were similar to those observed for speech/voice regions defined together (see Figure 7; blue and green dots). In both speech-sensitive and voice-sensitive regions, responses to low frequencies were significantly stronger than responses to middle and high frequencies (low frequency bias in 'speech vs. other':  $t = 5.59$ ,  $p < 0.01$ ; low frequency bias in 'voice vs. other':  $t = 6.09$ ,  $p < 0.01$ ).

		Comparison 1				Comparison 2				
		x	y	z	t	x	y	z	t1	t2
LH	Mid	-56	-19	5	3.2*	-56	-17	4	18.5**	10.2**
	Post	-54	-33	7	1.7	-56	-34	6	2.9*	2.5
RH	Ant	51	-1	-4	3.9*	51	0	-5	2.6	2.0
	Mid	55	-9	2	4.3*	56	-8	1	6.2**	7.1**
	Post	52	-31	3	3.7*	53	-31	3	14.4**	22.5**

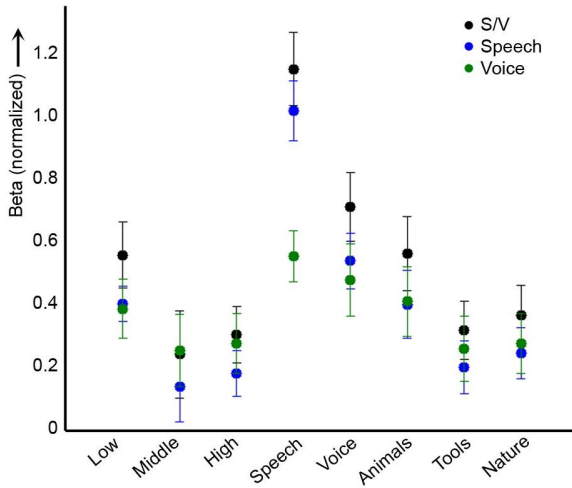
\* $p < 0.05$ ; \*\* $p < 0.01$

**Table 1. Talairach coordinates (x,y,z) of speech/voice regions and low frequency bias**

In 'Comparison 1', speech/voice regions are defined by the natural sounds experiment and quantification of the low frequency bias is performed on the localizer. In 'Comparison 2', speech/voice regions are defined by the localizer, and quantification of the low frequency bias is performed on the natural sounds experiment. Here, t1 indicates results based on all sounds, and t2 indicates results based on a subset of sounds with frequency content matched across sound categories. Locations of the speech/voice regions are reported as averaged Talairach coordinates [x, y, z]. Reported t-values result from paired t-tests contrasting the response to low frequencies vs. the response to higher frequencies.

Speech/voice regions, as defined by the localizer experiment, occupied tonotopic locations with preference for the lower range of the frequency scale (red-yellow colours, CF < 1 kHz; see Figure 6E, and column 2 in Figure 3 for single subject results). This preference was preserved when considering tonotopic maps obtained with a subset of sounds with frequency content matched across sound categories (see Methods, and Figure 6F). The analysis of regionally-averaged spectral profiles confirmed these results. For each speech/voice region, the main peak of the spectral profile was tuned to low frequencies (< 0.5 kHz). The bias was significant both when spectral profiles were estimated using all sounds (t1 in Table 1) and when a subset of sounds with frequency content matched across sound categories was used (t2 in Table 1). We observed a

significant low-frequency bias in all regions except posterior STG/STS in the left and anterior STG/STS in the right hemisphere (Table 1).



**Figure 7. Frequency bias in speech/voice regions**

The normalized response (beta values) of speech/voice regions (black), speech sensitive regions (blue), and voice sensitive regions (green; 300 most selective voxels as defined by the natural sounds experiment for each subject) to the tones and natural sounds in the localizer. Whiskers indicate standard error across subjects. Note that in all regions, responses are stronger to low (0.5 kHz) than to middle and high tones (1.5 and 2.5 kHz respectively). For quantification of this effect in speech/voice regions, see Table 1.

## Discussion

In this study, we employed a methodological approach that combines mathematical modeling of natural sounds with functional MRI to derive the spectral tuning curves of neuronal populations throughout the human auditory cortex. By extracting tonotopy and tuning width maps, we could depict the functional topography of the human auditory cortex. Moreover, we revealed an intrinsic relation between basic properties and categorical sensitivity in non-primary auditory cortex. That is, regions most responsive for speech and voice sounds have a bias towards low frequencies also when responding to non-preferred stimuli such as tones.

### Parcelation of the auditory cortex based on topographic maps

In agreement with previous fMRI studies (Formisano et al., 2003; Talavage et al., 2004; Da Costa et al., 2011; Striem-Amit et al., 2011; Langers & van Dijk, 2012), we observed several frequency selective clusters throughout the superior temporal cortex. To date, it remains unclear how the location and orientation of the auditory core relates to these tonotopic gradients. Several imaging studies suggested that the primary tonotopic gradient is oriented in postero-medial

to antero-lateral direction along HG (Seifritz et al., 2002; Formisano et al., 2003; Riecke et al., 2007). Conversely, recent studies argued that the main gradient runs in anterior-posterior direction (Humphries et al., 2010; Da Costa et al., 2011; Striem-Amit et al., 2011), possibly with a curvature between the two principal frequency gradients (Langers & van Dijk, 2012).

Here, we interpret maps of tonotopy and tuning width together. The maps of tuning width showed a region of narrow tuning flanked by regions of broader tuning along HG. This is consistent with the known organization of the monkey auditory cortex, where narrowly tuned primary areas are surrounded by a belt of more broadly tuned non-primary areas (Rauschecker et al., 1995; Hackett et al., 1998; Rauschecker & Tian, 2004; Kajikawa et al., 2005; Kusmirek & Rauschecker, 2009). Accordingly, we considered the narrowly tuned part of cortex in the HG region as the human PAC, oriented parallel to the main axis of HG. This region contains two mirror-symmetric frequency gradients (Figure 5C), which we interpret as representing two distinct cortical fields. These cortical fields may reflect the homologues of monkey primary fields AI and R (Kosaki et al., 1997; Hackett et al., 1998), and possibly correspond to the cyto-architectonically defined areas KAm and KAlt (Galaburda & Sanides, 1980) or region Te1 (Morosan et al., 2001). The additional frequency reversal observed in planum polare - anterior to the main tonotopic gradient – may correspond to the cyto-architectonically defined area PaAr (Galaburda & Sanides, 1980) and the human homologue of monkey RT (Kosaki et al., 1997; Hackett et al., 1998; see Figure 5C).

A precise delineation of functional areas outside the core regions remains difficult. Results from our field sign mapping suggested an additional gradient at the posterior-medial convergence of HG and HS (see Figure 5C, green region indicated with a triangle). This may correspond to cytoarchitectonically-defined area PaAc/d (Galaburda & Sanides, 1980) and the human homologue of monkey CM/CL (Kajikawa et al., 2005). Several additional gradients were present that paralleled the primary tonotopic gradient along HG in postero-medial to antero-lateral direction. Understanding the precise topology of the non-primary auditory areas defined by these gradients may require additional knowledge regarding the response properties (e.g. tuning to temporal/spectral modulations, latency) of these cortical regions. However, we suggest that the cortical region at the posterior-lateral adjacency of HG, separated from the core by its broader tuning width, might reflect the human homologue of the monkey belt areas ML (adjacent to hA1) and AL (situated adjacent to hR; Rauschecker & Tian, 2004). The location of this region is consistent with that of a cytoarchitecturally-defined subdivision of the human auditory cortex (i.e. area PaAi, Galaburda & Sanides, 1980; or Te2, Morosan et al., 2005).

### **Speech/voice-sensitive regions are embedded within topographic maps**

Beyond the Heschl's region, we observed clusters of frequency preference on PT and along STG and STS. Tonotopic maps in areas this far away from the primary auditory cortex have not been detected in animals and have remained largely elusive in previous human neuroimaging studies (note, however, a recent study by Striem-Amit et al., 2011). Stimulus choice may be highly relevant for extraction of tonotopic maps. A comparison between our tonotopic maps based on the natural sounds and tones experiment showed that especially outside the primary areas differences in estimated spectral tuning occurred. Previous studies often used artificial sounds, ranging from pure tones (Schonwiesner et al., 2002; Formisano et al., 2003; Langers et al., 2007; Woods et al., 2009; Da Costa et al., 2011), to frequency sweeps (Talavage et al., 2004; Striem-Amit et al., 2011), and alternating multi-tone sequences (Humphries et al., 2010). As natural sounds inherently engage auditory cortical neurons in meaningful and behaviorally relevant processing, they may be optimal for studying the functional architecture of higher order auditory areas.

Based on their anatomical location, we interpret the frequency selective clusters on PT and along STG as reflecting areas Te3, or PaAe/Tpt (Galaburda & Sanides, 1980; Morosan et al., 2005). Our tuning width maps show both narrow and broad tuning in these regions. The variations in tuning width on PT and along STG suggest that the multiple tonotopic maps in these regions encode sounds at different spectral resolutions. More specifically, the narrowly tuned regions on PT, and STG/STS may be involved in fine frequency discrimination required for processing spectrally complex sounds such as speech and music. So far, in animals, narrowly tuned regions have not been observed this far from the core. This discrepancy with our findings may reflect differences between species but may also be explained by methodological factors. fMRI typically has a much larger spatial coverage than microelectrode recordings, which could have caused inclusion of these regions in the current study where microelectrode studies failed to sample them. Alternatively, fMRI has a much poorer spatial resolution, and each voxel measures responses of a large group of individual neurons. A voxel containing neurons with diverse CF would be assigned with a broader width than a voxel containing neurons with more similar CF, which could have caused the appearance of narrowly tuned regions on the STG. Furthermore, our "tuning width" reflects the width of the main spectral peak only. As additional spectral peaks in the voxels' profiles are disregarded in the current study, the tuning width of a region is unrelated to its spectral complexity and thus also voxels with complex profiles can be labeled as narrowly tuned.



Overlaying cortical clusters with preference for speech and voice sounds on maps of tonotopy revealed an intrinsic relation between these two sound representations. That is, speech/voice sensitive regions displayed a bias towards low frequencies. This suggests that category-sensitive regions in the auditory cortex should not be interpreted as secluded modules of processing, but rather as part of large-scale topographic maps of sound features. Interestingly, this observed relation between topographic maps and category representations resembles the build-up of the visual system (Hasson et al., 2003; Rajimehr et al., 2011). For instance, in parallel with the embedding of speech/voice regions in the low-frequency part of a tonotopic map found in our study, face-selective areas in the occipito-temporal cortex occupy the foveal part of a single, unified eccentricity map (Hasson et al., 2003). Together, these results point to a general organizational principle of the sensory cortex, where the topography of higher order areas is influenced by their low-level properties.

We propose that our fMRI observations may reflect a neuronal filtering mechanism operating at the early stages of the transformation from tonotopic images of speech/voice sounds into their representations at a more abstract level. Speech and voice sounds contain most energy in relatively low frequency ranges ( $<1$  kHz; Crandall & MacKenzie, 1922) and this may be a distinctive feature with respect to other sound categories. Tuning of neuronal populations to the characteristic spectral components of a sound category (speech/voice) may thus act as a ‘matched-filter’ mechanism, which selectively amplifies category-characteristic spectral features and enhances the representational distance with respect to sounds from other categories. Our ‘matched filtering’ model agrees with the observation that closely matching the time-varying spectral characteristics between speech/voices and other sound categories reduces differences of related activations in STG/STS regions (Staeren et al., 2009). Furthermore, our findings bring support to recent psycholinguistic research that suggests a close link between general acoustic mechanisms and speech and voice perception. For instance, the frequency bias in the speech/voice sensitive regions predicts the psychophysical influences of the long-term average spectrum (LTAS) of non-speech stimuli on subsequent processing and perception of speech/voice (see e.g. Sjerps et al., 2011; Laing et al., 2012).

This observed low-frequency bias is most likely related to the input stage of speech/voice processing, and thus only partially explains the category-selective responses consistently observed in human and non-human auditory regions (Belin et al., 2000; Petkov et al., 2008). Understanding the entire computational chain underlying the nature of these responses remains an intriguing experimental challenge for future research.

## References

- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403: 309-312.
- Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I (2008) Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451: 197-201.
- Chi T, Ru P, Shamma SA (2005) Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am* 118: 887-906.
- Chueng SW, Bedenbaugh PH, Nagarajan SS, Schreiner CE (2001) Functional organization of squirrel monkey primary auditory cortex. *J Neurophysiol* 85: 1732-1749.
- Crandall IB, MacKenzie D (1922) Analysis of the energy distribution in speech. *Phys. Rev.* 19: 221-232.
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci* 31: 12067-14075.
- David SV, Mesgarani N, Fritz JB, Shamma SA (2009) Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli. *J Neurosci* 29: 3374-3386.
- Elhilali M, Shamma SA (2008) A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. *J Acoust Soc Am* 124: 3751-3771.
- Formisano E, Kim D, Di Salle F, van de Moortele PF, Ugurbil K, et al. (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40: 859-869.
- Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science* 322: 970-973.
- Friston KJ, Frith CD, Turner R, Frackowiak RS (1995) Characterizing evoked hemodynamics with fMRI. *Neuroimage* 2: 157-165.
- Galaburda A, Sanides F (1980) Cytoarchitectonic organization of the human auditory cortex. *The Journal of Comparative Neurology* 190: 597-610.
- Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp* 27: 392-401.
- Hackett TA, Stepniewska I, Kaas JH (1998) Subdivision of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394: 475-495.
- Hasson U, Harel M, Levy I, Malach R (2003) Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* 37: 1027-1041.
- Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. *Neuroimage* 50: 1202-1211.
- Imaizumi K, Priebe NJ, Crum PAC, Bedenbaugh PH, Chueng SW, Schreiner CE (2004) Modular functional organization of cat anterior auditory field. *J Neurophysiol.* 92: 444-457.
- Kajikawa Y, de la Mothe L, Blumell S, Hackett TA (2005) A comparison of neuron response properties in areas A1 and CM of the marmoset monkey auditory cortex: tones and broadband noise. *J Neurophysiol* 93: 22-34.
- Kay KN, David SV, Prenger RJ, Hansen KA, Gallant JL (2008a) Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. *Hum Brain Mapp* 29: 142-156.
- Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008b) Identifying natural images from human brain activity. *Nature* 452: 352-356.
- King AJ, Nelken I (2009) Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nature Neuroscience* 12: 698-701.
- Kosaki H, Hashikawa T, He J, Jones EG (1997) Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. *J Comp Neurol* 386: 304-316.
- Kusmierek P, Rauschecker JP (2009) Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J Neurophysiol* 102: 1606-1622.

- Laing EJ, Liu R, Lotto AJ, Holt LL (2012) Tuned with a tune: Talker normalization via general auditory processes. *Front Psychol*, doi: 10.3389/fpsyg.2012.00203.
- Langers DR, Backes WH, van Dijk P (2007) Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. *Neuroimage* 34: 264-273.
- Langers DR, van Dijk P (2012) Mapping the tonotopic organization of the human auditory cortex with minimally salient acoustic stimulation. *Cerebral Cortex*, doi: 10.1093/cercor/bhr282.
- Merzenich MM, Brugge JF (1973) Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res* 50: 275-296.
- Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, et al. (2001) Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13: 684-701.
- Morosan P, Schleicher A, Amunts K, Zilles K (2005) Multimodal architectonic mapping of human superior temporal gyrus. *Anat Embryol* 210: 401-406.
- Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. *Neuron* 63: 902-915.
- Naselaris T, Kay KN, Nishimoto S, Gallant JL (2010) Encoding and decoding in fMRI. *Neuroimage* 56: 400-410.
- Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF (2012) Reconstructing speech from human auditory cortex. *PLoS Biol* 10: e1001251. doi:10.1371/journal.pbio.1001251.
- Perrodin C, Kayser C, Logothetis NK, Petkov CI (2011) Voice cells in the primate temporal lobe. *Current Biology* 21: 1408-1415.
- Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008) A voice region in the monkey brain. *Nat Neurosci* 11: 367-374.
- Petkov CI, Logothetis NK, Obleser J (2009) Where are the human speech and voice regions, and do other animals have anything like them? *Neuroscientist* 15, 419-429.
- Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RBH (2011) The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biology* 9: doi:10.1371/journal.pbio.1000608.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111-114.
- Rauschecker JP, Tian B (2004) Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91: 2578-2589.
- Riecke L, van Opstal AJ, Goebel R, Formisano E (2007) Hearing illusory sounds in noise: Sensory-perceptual transformation in primary auditory cortex. *J Neurosci* 27: 12684-12689.
- Schonwiesner M, von Cramon DY, Rubsamen R (2002) Is it tonotopy after all? *Neuroimage* 17: 1144-1161.
- Seifritz E, Di Salle F, Esposito F, Herdener M, Neuhoﬀ JG, Scheﬄer K (2006) Enhancing BOLD response in the auditory system by neurophysiologically tuned fMRI sequence. *NeuroImage* 29: 1013-1022.
- Sjerps MJ, Mitterer H, McQueen JM (2011) Constraints on the processes responsible for the extrinsic normalization of vowels. *Atten Percept Psychophys* 73: 1195-215.
- Staeren N, Renvall H, De Martino F, Goebel R, Formisano E (2009) Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology* 19: 498-502.
- Striem-Amit E, Hertz U, Amedi A (2011) Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLoS One* 6: DOI:10.1371/journal.pone.0017832.
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Stuttgart: G. Thieme.

- Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, et al. (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol* 91: 1282-1296.
- Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315-2331.
- Woods DL, Stecker GC, Rinne T, Herron TJ, Cate AD, et al. (2009) Functional maps of human auditory cortex: effects of acoustic features and attention. *PLoS One* 4: doi:10.1371/journal.pone.0005183.
- Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: Music and speech. *Trends Cogn Sci* 6: 37-46.

### **Acknowledgements**

We thank R. Goebel, P. de Weerd, and G. Valente for comments and discussions. We thank R. Santoro for implementation of the early auditory processing model. This work was supported by Maastricht University and the Netherlands Organization for Scientific Research (NWO), Top talent grant 021-002-102 (M.M.) and Innovational Research Incentives Scheme Vidi grant 452-04-330 (E.F.).



## **Chapter 3**

### **Spatial organization of frequency preference and selectivity in the human inferior colliculus**

\*Authors share equal contribution

De Martino F\*, Moerel M\*, van de Moortele PF, Ugurbil K, Goebel R, Yacoub E, Formisano E (2013) Spatial organization of frequency preference and selectivity in the human inferior colliculus. Nat Comm 22, doi: 10.1038/ncomms2379.



## Abstract

To date, the functional organization of human auditory sub-cortical structures can only be inferred from animal models. Here, we use high-resolution functional MRI at ultra-high magnetic fields (7 Tesla) to map the organization of spectral responses in the human inferior colliculus (hIC), a sub-cortical structure fundamental for sound processing. We reveal a tonotopic map with a spatial gradient of preferred frequencies approximately oriented from dorso-lateral (low frequencies) to ventro-medial (high frequencies) locations. Furthermore, we observe a spatial organization of spectral selectivity (tuning) of fMRI responses in the hIC. Along isofrequency contours, fMRI-tuning is narrowest in central locations and broadest in the surrounding regions. Finally, by comparing sub-cortical and cortical auditory areas we show that fMRI-tuning is narrower in hIC than on the cortical surface. Our findings pave the way to non-invasive investigations of sound processing in human sub-cortical nuclei and to studying the interplay between sub-cortical and cortical neuronal populations.



## Introduction

The inferior colliculus (IC) – centrally located in the auditory pathway – is an obligatory relay station for all information ascending from brainstem nuclei to thalamus and cortex (Winer and Schreiner, 2005). The IC is pivotal for a vast number of auditory tasks (Aitkin and Martin, 1987; Ehret and Merzenich, 1988; Schreiner and Langer, 1997; Versnel et al., 2009) and thus, understanding its functional organization is key in advancing knowledge on how sound perception emerges in the human brain.

The functional organization of the IC has been investigated in a variety of non-human species (e.g. rats, cats, squirrel monkey) using invasive electrophysiological recordings (Merzenich and Reid, 1974; Schreiner and Langner, 1988; Winer and Schreiner, 2005; Malmierca et al., 2008) and, more recently, non-invasively using functional MRI (fMRI; Baumann et al., 2011; Cheung et al., 2012). Various studies showed that the IC is tonotopically organized (Winer and Schreiner, 2005). In the rat and in the cat, the orientation of the main tonotopic gradient (low to high frequency preference) is orthogonal to the fibrodendritic laminae in the central nucleus of the IC (Malmierca et al., 1995; 2008) and runs in the dorso-lateral to ventro-medial direction (Merzenich and Reid, 1974). Based on these results in animals and on the observation of tonotopic maps in the human cortex (Formisano et al., 2003; Talavage et al., 2004; Humphries et al., 2010; Woods et al., 2010; Da Costa et al., 2011; Langers and van Dijk, 2011; Striem-Amit et al., 2011), the human IC (hIC) is assumed to be tonotopically organized as well. However, there is no experimental evidence for the existence and spatial layout of a tonotopic map of hIC. Furthermore, in humans, it is unknown whether selectivity of spectral tuning (i.e. tuning width) is homogeneous or location dependent throughout the hIC, a question so far addressed only by studies in the cat IC (Schreiner and Langner, 1988). Finally, while electrophysiological recordings revealed fine frequency tuning in the human and animal auditory cortex (Bitterman et al., 2008), no information is available to compare tuning width between the human midbrain and cortex in response to the same stimuli.

In the present study, we examined the spatial organization of spectral responses in hIC using fMRI at 7 Tesla. Our analysis of fMRI responses to simple tones and natural sounds revealed the detailed layout of the tonotopic organization within hIC. Furthermore, the estimation of voxels' spectral response profiles (SRP) with mathematical modelling of the fMRI responses to the natural sounds (i.e. fMRI encoding; Kay et al., 2008; Mitchell et al., 2008; Naselaris et al., 2009; 2010) enabled us to map the spectral selectivity of different locations in the

human IC (fMRI tuning) and compare it to that of primary and non-primary auditory cortical areas.

## Methods

### Subjects

The study consisted of three experiments. Five subjects (median age = 32, 3 males) participated in both Experiment 1 and 2. Four subjects (median age = 30; 3 males) participated in Experiment 3. The subjects had no history of hearing disorder or neurological disease. The imaging protocol used in this study was approved by the Institutional Review Board (IRB) of the University of Minnesota.

### Stimuli

In Experiment 1, 800 ms long amplitude-modulated (8 Hz; modulation depth of 1) tones (.45; .5; .55; 1.35; 1.5; 1.65; 2.25; 2.5; 2.75 kHz) were presented grouped in blocks around three centre frequencies (.5; 1.5; 2.5 kHz) together with blocks of natural sounds (800 ms recordings) grouped in blocks of five different categories (speech, voice, animals, tools, nature). In Experiment 3, 800 ms long amplitude-modulated (8 Hz; modulation depth of 1) tones (.16; .18; .19; .28; .30; .32; .48; .51; .55; .81; .86; .93; 1.37; 1.46; 1.57; 2.31; 2.48; 2.66; 3.91; 4.19; 4.49; 6.61; 7.09; 7.60 kHz) were presented grouped in blocks around eight centre frequencies (.18; .30; .51; .86; 1.46; 2.48; 4.19; 7.09 KHz). In both Experiment 1 and 3, each block consisted of six sounds, lasted 18 seconds and was separated by 12 seconds of no stimulation from the following block. Each condition was presented twice per run, and we acquired six runs per subject.

In Experiment 2, stimuli were derived from recordings of natural sounds ( $n = 144$ ; e.g. speech sounds, vocalizations, animal cries, tools, nature sounds), and cut to a one second duration. Each stimulus was presented three times across six different runs, and the inter-stimulus-interval (ISI) was jittered between 5200 and 10400 ms. Zero trials (where no sound was presented, 6% of total trials) and catch trials (repeating the sound of the previous trial, 5% of total trials) were added. Subjects were asked to press a button when a sound was repeated; these catch trials were excluded from the analysis of the data. In order to ensure a clear perception of the sounds during scanning, in all experiments sounds were presented in silent gaps in between the acquisitions of two subse-

quent brain volumes (van Atteveldt et al., 2004; Riecke et al., 2007; Formisano et al., 2008).

In all three experiments, sounds were sampled at 44.1 kHz. Sound onset and offset were ramped with a 10 ms linear slope and sound energy (RMS) was equalized. Prior to scanning, all sounds were subjectively equalized for loudness. In Experiments 1 and 2, sounds were presented with an MR compatible audio system based on air tubes (Avotec Inc.) with a linear frequency transfer function up to about 4 kHz. The system was modified in such a way that the final piece of the air tubes was directly inserted in the subjects' ear through mouldable earplugs. In Experiment 3, sounds were presented with an MR compatible piezoelectric/ear bud system (Sensimetric Inc.) with a linear frequency transfer function up to about 8 kHz. In all experiments sounds were presented at about 60 dB. All subjects were asked to cover their ears with memory foam pieces cut to an average ear size for additional sound protection.

### Imaging parameters

Data were acquired on a 7T whole body system driven by a Siemens console using a head gradient insert operating at up to 80 mT/m with a slew rate of 333 T/m/s. A head RF coil (single transmit, 16 receive channels) was used to acquire anatomical ( $T_1$  weighted) and functional ( $T_2^*$  weighted BOLD) images. In each experiment,  $T_1$  weighted ( $1 \text{ mm}^3$ ) images were acquired using a modified Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) sequence (TR=2500 ms; TI=1500 ms; TE=3.67 ms). Proton density (PD) images were also acquired together with the  $T_1$  weighted images (both acquisitions are interleaved in the modified MPRAGE sequence) and were used to minimize inhomogeneities in  $T_1$  weighted images (van de Moortele et al., 2009). Acquisition time for anatomical images was ~ 7 minutes.  $T_2^*$  weighted functional data were acquired using a clustered Echo Planar Imaging (EPI) technique. In Experiment 1 and 3, each brain volume was scanned at 1.5 mm isotropic resolution (44 slices; GRAPPA acceleration X3; partial Fourier 6/8; TR = 3000 ms; time of acquisition [TA] = 1500 ms). In Experiment 2, brain volumes were scanned at 1.5 mm isotropic resolution (31 slices; GRAPPA acceleration X3; partial Fourier 6/8; TR = 2600 ms; time of acquisition [TA] = 1200 ms). For all experiments, slices were placed in order to maximize efficiency (short acquisition time) and cover the brain transversally from the inferior portion of the anterior temporal pole to the superior portion of the superior temporal gyrus bilaterally. All functional sessions consisted of six runs of approximately 10 minutes each. As cardiac and respiration cycles were not recorded, we did not apply any post-hoc correction to reduce cardiac induced pulsation noise that may decrease the

temporal SNR in brainstem regions. Despite this we found reliable activation in the hIC to all presented sounds.

Functional and anatomical data were analyzed with BrainVoyager QX. Pre-processing consisted of slice scan-time correction (with sinc interpolation), temporal high pass filtering (removing drifts of 3 cycles or less per run in data collected from Experiments 1 and 3 and of 7 cycles or less per run in data collected from Experiment 2) and 3-dimensional motion correction. Additional spatial smoothing (Gaussian kernel with FWHM = 2 mm) and temporal smoothing (2 consecutive data points) were applied to the data collected in Experiments 1 and 3. Functional data were co-registered to the anatomical data collected in the same experiment. Anatomical and functional data were co-registered across sessions and normalized in Talairach space. Functional data were resampled (with sinc interpolation) in the normalized space at a resolution of 0.5 mm isotropic and, for the hIC analysis, limited to a bounding box (60 x 70 x 60 mm) surrounding the inferior colliculus bilaterally.

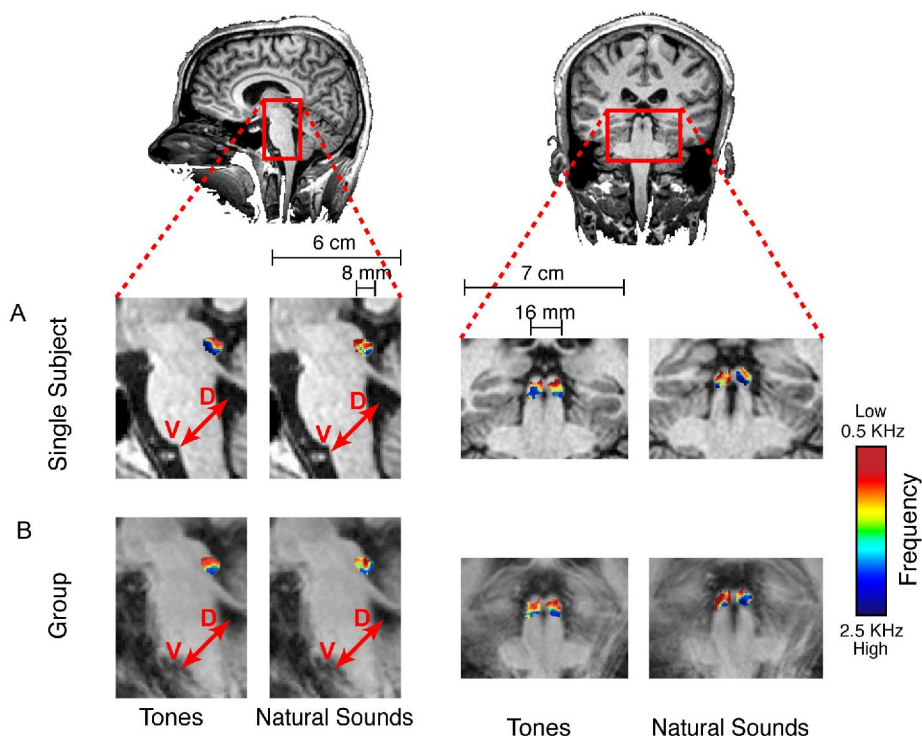
### Computation of topographic maps

The response to simple tones and natural sounds was evaluated by computing a fixed effects General Linear Model (GLM) of data collected in Experiments 1 and 3. A region of interest was created based on these activation maps ( $Q[FDR] < 0.001$ ) and was used in all subsequent analyses. Additionally, we used an anatomical constraint for the sub-cortical analysis, by considering gray/white matter voxels only (i.e. intensity threshold  $> 100$  in the mean anatomical images).

The spatial layout of frequency selective responses elicited by simple tones (i.e. Experiments 1 and 3) was obtained by computing “best frequency” maps (Formisano et al., 2003). A single subject GLM analysis was used to compute the response to the three centre frequencies (.5; .1.5; 2.5 kHz) in Experiment 1 and to the eight centre frequencies (.18; .30; .51; .86; 1.46; 2.48; 4.19; 7.09 kHz) in Experiment 3. The voxels’ best frequency (BF) was assigned for each run separately as the frequency to which the voxel responded strongest. A leave-one-run out cross-validation procedure was used to assess the reproducibility of these maps within each single subject. Individual subject tonotopy maps were created by averaging the voxels’ BF across runs, and by colour-coding the resulting map according to a linear red-yellow-green-blue colour scale. Group maps were obtained by averaging single subject BF maps.

Tonotopic and tuning width maps from Experiment 2 were obtained using an “encoding” technique (Kay et al., 2008; Mitchell et al., 2008; Naselaris et al., 2009; 2010) that allows estimating the voxels’ response to features (e.g. fre-

quency; voxels' SRP) from the fMRI patterns elicited by complex sounds. For details on the encoding technique, see chapter 2. Compared to the procedure performed in chapter 2, two details in the analysis were changed. First, the frequency content of all natural sounds ( $N = 144$ ) was characterized by means of a Fourier analysis. Second, a 20-fold cross-validation procedure was used to estimate matrix each voxels' SRP using a regularized regression approach (RVM), leaving out 24 sounds per fold.



**Figure 1. Tonotopic maps in the inferior colliculus**

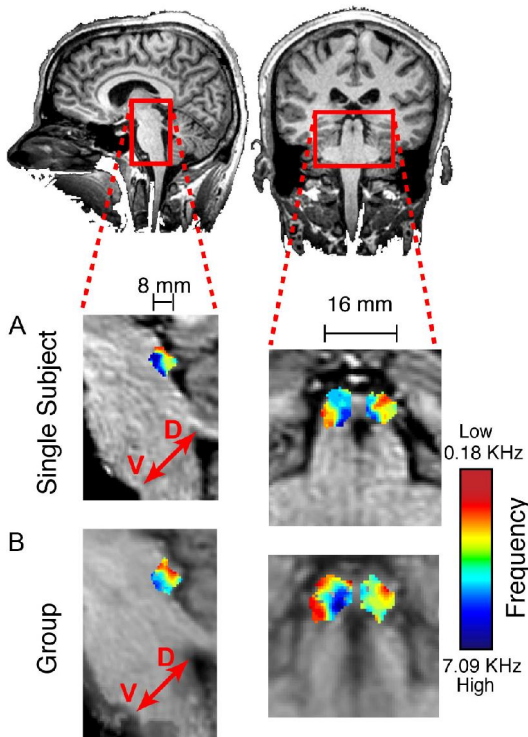
Tonotopic maps are shown (A) for a representative subject and (B) in the group, obtained from single tones and natural sounds (Experiment 1-2). Maps are superimposed on high resolution anatomical images of the subject (A) and to the anatomical images obtained from the average of all subjects (B). The red arrow in the sagittal sections indicates the dorsal (D) to ventral (V) direction in hIC.

For each fold, best frequency (BF) and tuning width (TW) were estimated per voxel using a two-step procedure. First, we computed the average voxel response within three frequency windows (centre: .5; .1.5; 2.5 KHz; width 600 Hz). The centre frequency (in Hz) of the maximum responding window was

assigned as the voxels' BF. Tonotopic maps were obtained by colour-coding the estimated BF using a linear red-yellow-green-blue colour scale. Second, we fitted a Gaussian function to the full voxels' SRP with mean constrained within the frequency window eliciting the most consistent maximum response across all folds. For each voxel  $j$ , the full width at half maximum (FWHM) of the fitted Gaussian in Hz was used to compute the tuning width (TW) as:

$$TW_j = CF_j / FWHM_j$$

We assigned significance to the obtained BF and TW values by means of a permutation analysis as in chapter 2 ( $N = 500$  per subject). We observed, in correspondence with previous studies in a variety of animals, a positive correlation between maps of TW and BF (Cheung et al., 2001; Read et al., 2001). The dependence of TW on BF was estimated with a smoothing spline fit. The residuals of this fit ( $TW_{res}$ ) were used to create final maps of tuning width. Tuning width maps were obtained by colour-coding the estimated  $TW_{res}$  using a linear yellow-green-blue-purple colour scale. Group maps were obtained by averaging single subject BF and  $TW_{res}$  maps.



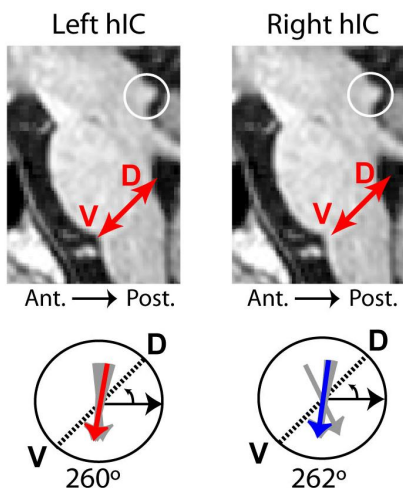
**Figure 2. Tonotopic maps obtained from single tones**

Tonotopic maps are shown (A) for a representative subject and (B) in the group, obtained from single tones (eight center frequencies, Experiment 3). Sagittal and coronal views of the maps are superimposed on high resolution anatomical images of the subject (A) and to the anatomical images obtained from the average of all subjects (B). The red arrow in the sagittal sections indicates the dorsal (D) to ventral (V) direction in the human IC.

### Evaluation of tonotopy and tuning width

The correspondence between tonotopic maps obtained from simple tones (Experiment 1) and natural sounds (Experiment 2) was evaluated by computing the correlation between these two tonotopic maps in each subject and between group maps. Significance was obtained with permutation testing ( $N = 500$ ) while accounting for the spatial correlation present in the single maps (i.e. by applying the same smoothing procedure to the unpermuted and permuted maps).

In order to quantify the direction of the main frequency gradient, a gradient analysis was performed on the tonotopic maps. For each voxel, we computed - in subsequent sagittal slices - the direction of increasing best frequency with respect to a transversal (i.e. horizontal) vector pointing in the anterior to posterior direction. To avoid nuisance effect from voxels lying at the border of the map, the direction was averaged across voxels within a region of interest obtained by eroding the original hIC mask (i.e. eliminating the outer most shell of the hIC). This analysis was performed both at individual subject level (for maps resulting from Experiment 1 only) and in the group maps (for maps resulting from Experiments 1-3).



**Figure 3. Quantification of tonotopic gradient direction**

Direction of the frequency gradient in the left and right human inferior colliculus (IC) as obtained from the response to simple tones. Angles are referred to a vector pointing in the anterior to posterior directions as indicated in the figure. Gray arrows indicate single subject results; colored arrows indicated results obtained from the group tonotopic maps resulting from Experiment 1. The dorsal (D) to ventral (V) direction in the human IC is indicated by a red line on the anatomy, and by a black dotted line in the vector plots.

We further examined the pattern of tuning width in the hIC for individual subjects and the group map, by performing an additional analysis on each of these maps. First, for each tonotopic map as extracted in Experiment 1 we selected eight equally spaced sagittal slices (four slices in left and right hemisphere each; slices selected in centre of hIC to minimize the variation in size of the maps). Second, we defined the iso-frequency contour at 1.5 kHz (i.e. middle frequency) in each slice. Third, all iso-contours were linearly interpolated to

hundred dimensional vectors and aligned to each other. The average length of these iso-contours was approximately 8 mm. Finally, we sampled the tuning width map as extracted from Experiment 2 along each contour.

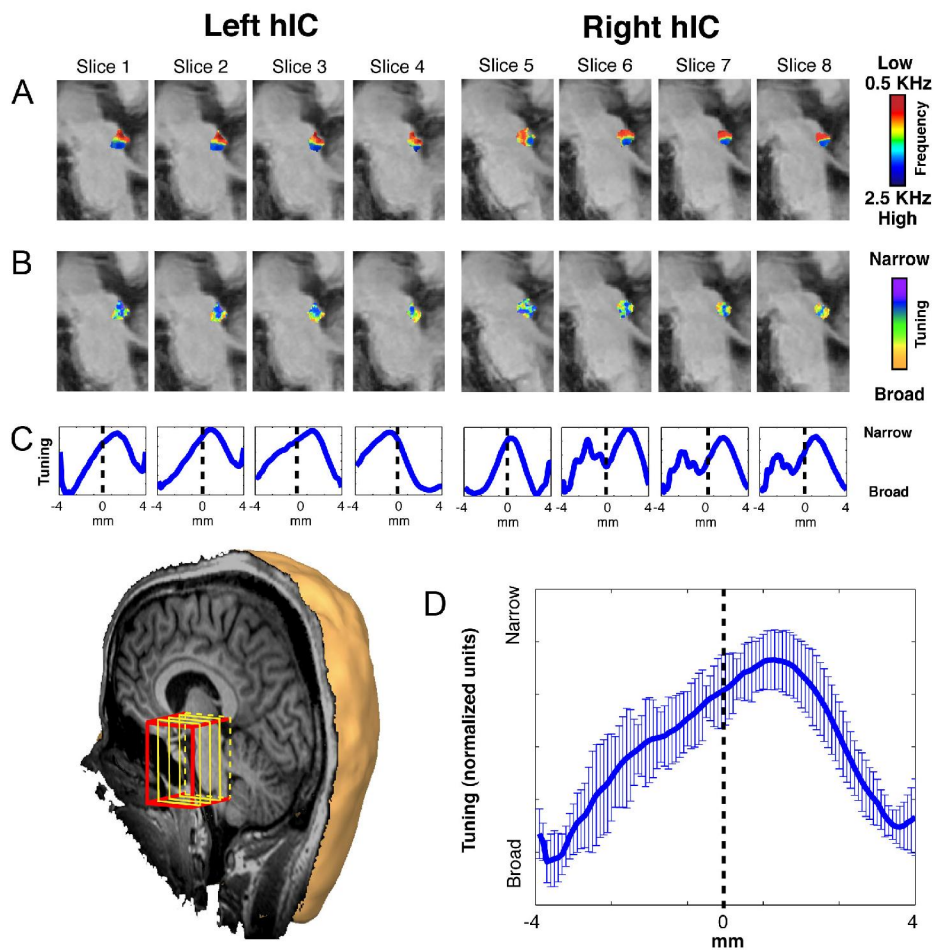
Finally, to examine whether the spatial distribution of tuning width in hIC could result from vasculature differences throughout hIC, we evaluated the correlation between tuning width and the overall response to natural sounds (F-Map; data from Experiment 2) or to the beta-map representing the response amplitude to natural sounds.

### **Parcelation of cortical auditory fields**

We delineated three cortical auditory fields (i.e. hA1, hR, and lateral belt) based on the combination of group tonotopic and tuning width maps. Based on observations from monkey auditory cortex (Rauschecker et al., 1995; Hackett et al., 1998; Rauschecker and Tian, 2004; Kajikawa et al., 2005; Kusmirek and Rauschecker, 2009) and recent findings in the human auditory cortex (Moerel et al., 2012; see chapter 2), we defined the narrow region along HG as the auditory core. This region was divided into a medial and lateral auditory at the low frequency gradient reversal (Hackett et al., 2001). The resulting regions each contained a complete frequency gradient and may reflect the human homologues of monkey A1 and R (Merzenich and Brugge, 1973; Hackett et al., 2001). Laterally and immediately adjacent to the auditory core, occupying Heschl's sulcus (HS) and the anterior part of planum temporale, we identified an auditory belt area. This area was distinguished from the auditory core by its tuning width, which was consistently broader (Rauschecker et al., 1995; Hackett et al., 1998; Rauschecker and Tian, 2004; Kajikawa et al., 2005; Kusmirek and Rauschecker, 2009).

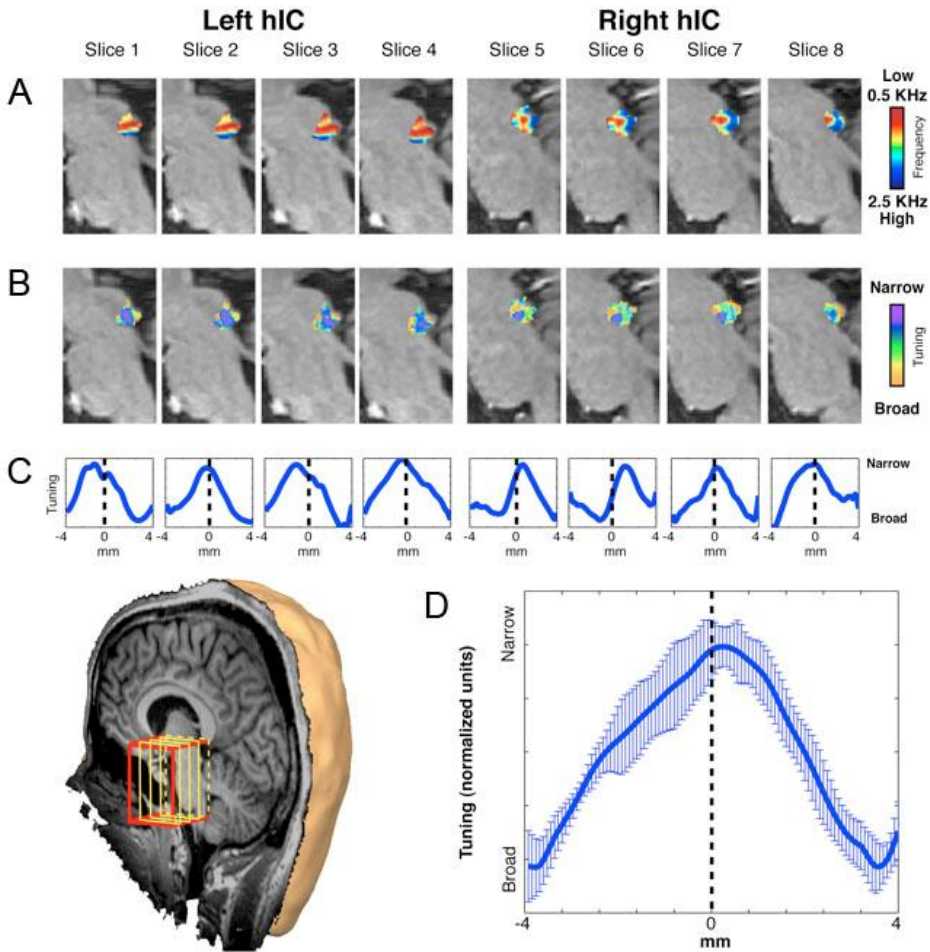
The cortical regions of interest - as drawn on the average cortical surface of all five subjects - were projected back to the individual volumes, and for each region the distribution of tuning width (TW) was sampled. All distributions were normalized for the total number of voxels in each area. Note that, as  $TW_{res}$  values were computed by removing the relation between TW and BF separately for cortical and subcortical regions, we could not use these values for comparisons across regions. Instead, we used uncorrected TW for this part of the analysis. We tested, in four intervals equally spaced between the minimum and maximum observed tuning widths, for differences in number of voxels across the four regions. We also tested (non parametric test Wilcoxon rank-sum test) for differences in mean tuning between areas and across subjects.





**Figure 4. Tuning width**

(A) Eight equally spaced sagittal slices (four in the left and right IC) of the group tonotopic map as obtained in Exp. 1 ('tones') superimposed to group-averaged anatomy. (B) Group tuning width map as obtained in Experiment 2 ('natural sounds'). (C) Spectral tuning sampled along the 1.5 kHz iso-frequency line for the same slices presented in A-B (D) Averaged distribution (mean and standard deviation across slices) of spectral tuning along the 1.5 kHz iso-frequency line. In C-D, the dashed line marks the center of the iso-frequency line.



**Figure 5. Tuning width in a representative subject.**

(A) Functional maps (single subject) of spectral preference (i.e. tonotopy; ‘tones’) in hIC (eight sagittal slices superimposed on subjects’ anatomical data). (B) Spectral tuning (i.e. tuning width; ‘natural sounds’), as obtained in the same subject. (C) Sampling of the spectral tuning width map along the selected (1.5 kHz) iso-frequency line for the eight slices separately. (D) Average distribution across slices (mean and standard deviation) of spectral tuning along the iso-frequency line. Dashed lines in C-D mark the center of the iso-frequency line.

## Results

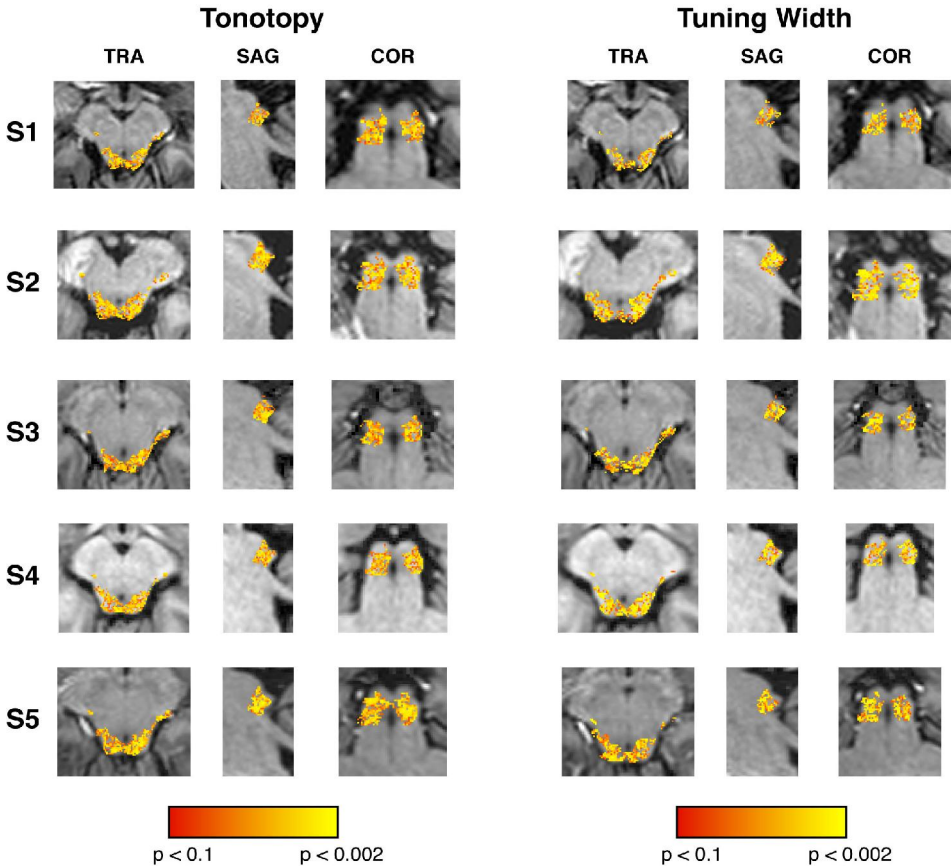
### FMRI best-frequency maps in the inferior colliculus

In all subjects, we observed reliable bilateral activation of the IC (Guimaraes et al., 1998; Baumann et al., 2010; 2011). The medial geniculate nucleus of the thalamus (MGN) was less consistently activated (in two of subjects bilaterally, and two subjects only unilaterally) and therefore was excluded from further analysis. Furthermore, we observed consistent activation in the auditory cortex, including the Heschl's region (approximately corresponding to the primary auditory core; Formisano et al., 2003), and surrounding regions in the planum temporale (PT) and STG. Figures 1 and 2 display the frequency preference (tonotopic maps) as extracted by Experiments 1-2 and Experiment 3 respectively. Although differences can be observed, in maps from all three experiments the spatial distribution of best-frequency progressed approximately from dorsal-lateral locations (low frequencies, red colour) to ventral-medial locations (high frequencies, blue colour). This apparent consistency was confirmed by a significant positive spatial correlation in maps across Experiment 1 and 2 at the group level (over 3602 voxels;  $p < 0.001$ , permutation testing) and at single-subject level in four of five subjects ( $p < 0.05$ , permutation testing). Quantification of the direction of the main frequency gradient in the group maps showed that this gradient was oriented at an angle of  $260^\circ$  (left hemisphere) and  $262^\circ$  (right hemisphere) with respect to the horizontal axis in Experiment 1 (variability of  $\pm 10^\circ$  across sagittal slices; see Fig. 3), at an angle of  $252^\circ$  (left hemisphere) and  $261^\circ$  (right hemisphere) in Experiment 2 (with a variability of  $\pm 15^\circ$  across sagittal slices) and at an angle of  $262^\circ$  (left hemisphere) and  $265^\circ$  (right hemisphere) in Experiment 3 (with a variability of  $\pm 10^\circ$  across sagittal slices).

### Frequency tuning of fMRI responses in the inferior colliculus and cortex

In group maps and at the level of individual subjects, we observed a spatial pattern of tuning width in hIC (see Fig. 4-5 for group maps and results from a representative subject respectively, and Fig. 6 for the significance of extracted topographic maps). That is, we observed narrower tuning in central portions of the IC (blue-violet in Fig. 4B-5B) and broadly tuned voxels (yellow-orange in Fig. 4B-5B) on its outer shell. This spatial distribution of fMRI tuning within the human IC was further examined by sampling the tuning width along the iso-frequency contour of the middle frequency (1.5 kHz) region in the tonotopic maps (Fig. 4C-5C). Both in single sagittal slices and in mean across the human IC (Fig. 4D-5D), tuning width was narrower at the centre of the iso-frequency

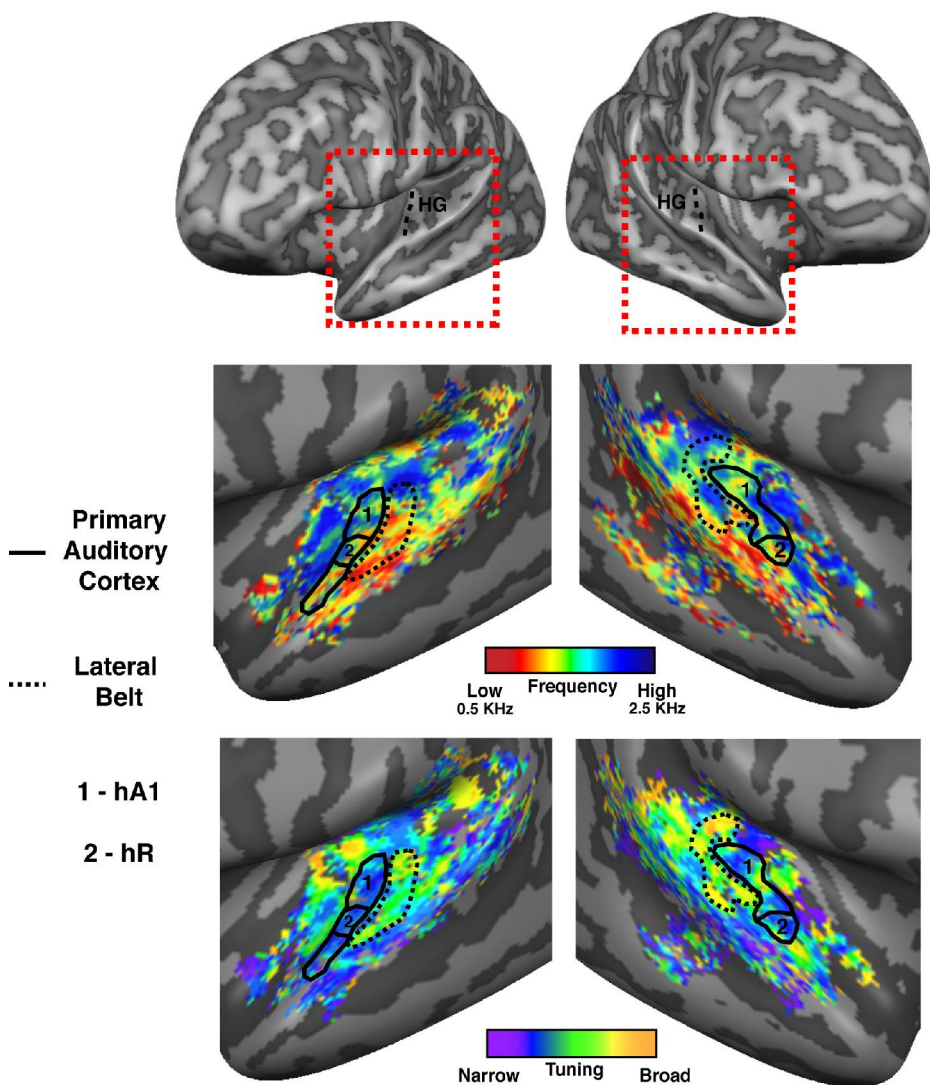
contour (high values of  $TW_{res}$ ), and broadened at its outer edges (low values of  $TW_{res}$ ). As we observed no correlation between tuning width and the overall response to natural sounds, this spatial distribution could not be explained by variations in overall activation levels of IC voxels.



**Figure 6. Permutation analysis to evaluate significance**

Results of the single subject permutation analysis ( $N = 500$ ) for the assessment of significance of center frequency (left column) and tuning width (right column) in the natural sounds modeling. Individual maps are thresholded at  $p < 0.01$  uncorrected.

Next, we defined cortical regions hA1, hR, and lateral belt (Fig. 7), and compared at single subject level the proportion of narrowly and broadly tuned voxels within the hIC to that of auditory cortical regions hA1, hR, and lateral belt (see Methods; Fig. 8). At the single subject level, we observed a wide distribution of TW values. However, a larger proportion of voxels with narrower tuning was present in the IC (gray curves in Fig. 8) compared with all other cortical areas.

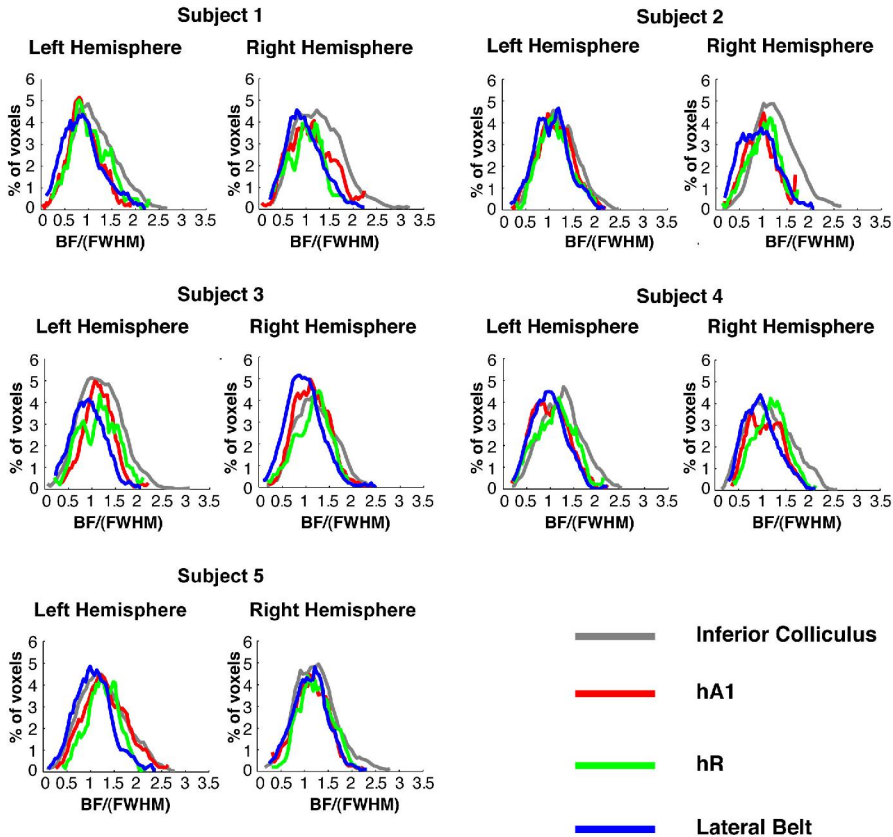


**Figure 7. Parcelation of superior temporal plane**

Parcelation of auditory cortical areas based on group tonotopy and tuning width maps. Group functional maps from the natural sounds experiment of spectral preference (i.e. tonotopy; second row) and selectivity (i.e. tuning width; third row) are sampled on the average (after cortex based alignment) cortical surface of the five subjects. In the second and third row, full and dashed lines outline the distinction between primary auditory cortex and lateral belt areas. Numbering within the primary auditory cortex highlights the separation between hA1 and hR.

While in most subjects the tuning width in cortical regions was limited to 0.6 octaves (TW = 2.5) or broader, in the IC several voxels were more narrowly tuned (up to 0.45 octaves; TW = 3.23). We observed significant differences

between lateral belt and hIC in the narrowest and broadest tuning width bin (see Fig. 9). The lateral belt contained more broadly tuned voxels than hIC ( $p < 0.05$ ), while hIC contained more narrowly tuned voxels than lateral belt ( $p < 0.05$ ). When testing for differences in mean tuning width between areas across subjects (Wilcoxon rank-sum test), we observed narrower fMRI tuning in the IC compared to hR and lateral belt areas bilaterally ( $p < 0.05$ ), and narrower fMRI tuning in right IC than in right hA1 ( $p < 0.05$ ).



**Figure 8. Visualization of tuning width across regions**

Histograms of voxels' tuning width (BF in Hz/FWHM in Hz) for each region (inferior colliculus – gray, hA1 – red, hR – green, lateral belt – blue) and hemisphere. Note that a larger proportion of narrowly tuned voxels is present in hIC (gray curves; high values), while lateral belt regions contain a larger proportion of broadly tuned voxels (blue curves; low values).



## Discussion

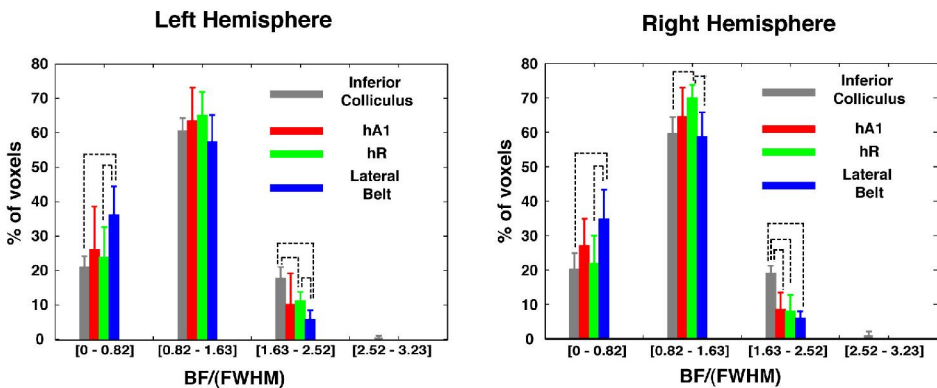
In this study, we used high-resolution fMRI to map the spatial organization of spectral responses in the human inferior colliculus based on responses to simple tones (Experiment 1 and 3) and natural sounds (Experiment 2). Across subjects and experiments, we consistently observed a single low-to-high tonotopic gradient in hIC, approximately oriented along a dorso-lateral to ventro-medial direction. Furthermore, spectral selectivity (fMRI tuning) based on responses to natural sounds was organized in a narrow-to-broad tuning pattern in hIC when progressing from central to peripheral locations. Finally, we showed that fMRI tuning in hIC was narrower than in primary (right hemisphere) and non-primary auditory (bilaterally) cortical areas. Our findings enable a comparison of human and animal auditory systems at the sub-cortical processing level and pave the way to the mapping of relevant acoustic features in the human auditory sub-cortical pathway.

### Tonotopy in the human inferior colliculus

The presence of a tonotopic map in the hIC is consistent with early reports in animals (e.g. cat and monkey) obtained using invasive electrophysiological recording (Merzenich and Reid, 1974; FitzPatrick, 1975) and, more recently, with fMRI in the monkey (Baumann et al., 2011). The IC frequency gradient observed in non-human species was oriented in a dorso-lateral to ventro-medial direction (Merzenich and Reid, 1974; FitzPatrick, 1975; Schreiner and Langner, 1988; 1997; Baumann et al., 2011; Cheung et al., 2012). In hIC, the dorsal to ventral direction runs orthogonal to the direction of the brain stem, at an orientation of  $225^\circ$  relative to the anterior-posterior horizontal axis (i.e. red arrow in Fig. 1 and 2, and black dotted line in Fig. 3). Our analysis revealed a tonotopic gradient oriented at  $\sim 260^\circ$  with respect to the anterior to posterior direction, which represents a deviation of  $35^\circ$  with respect to the dorsal-to-ventral direction. It is challenging to say whether this difference reflects an anatomical shift of the colliculus with respect to the brain stem across species or simply results from the inaccurate alignment of previous results with the present maps. The comparison of imaging data across species (Petkov et al., 2006; Baumann et al., 2011) with careful realignment to the same reference system may enable a more precise comparison.

Furthermore, three-dimensional representations of the tonotopic gradient from electrophysiology studies in animals have shown a gradual tilt of the iso-frequency laminae in the rostral and lateral directions (Merzenich and Reid, 1974; Semple and Aitkin, 1979). Despite the relatively high spatial resolution of

our measurements, our spatial resolution does not allow drawing conclusions regarding such fine-grained deviations in orientation of the iso-frequency contours within hIC. Instead, our gradient analysis of best-frequency maps was limited to the main frequency gradient (i.e. averaged across all pixels in sagittal slices of the hIC). Voxels whose best frequency does not follow this principal gradient direction are visible in selected slices, particularly in the tonotopic maps obtained with natural sounds (Figure 1). We expect that these variations may reflect an irregular spatial distribution of the frequency gradient. Alternatively, they could result from the inaccurate estimation of the voxels' spectral response profiles.



**Figure 9. Differences in tuning width across regions**

The percentage of voxels with spectral tuning (BF/FWHM) within four different windows ([0 - 0.82], [0.82 - 1.63], [1.63 - 2.52], [2.52 - 3.23]) is reported for four different areas (inferior colliculus - gray, hA1 - red, hR - green, lateral belt - blue) in both left (left column) and right (right column) hemispheres. Error bars represent standard deviations across subjects. Dashed lines mark significant differences (Wilcoxon ranksum test;  $p < 0.05$ ).

### Spectral selectivity throughout the auditory system

Beyond maps of tonotopy, we revealed a spatial organization of tuning width of hemodynamic responses in hIC. That is, narrow tuning was observed in the centre of the hIC, and broad tuning at its extremes. Given the complexity of the fMRI signal, both vascular and neuronal factors may contribute to this observation. The distribution of fMRI tuning width values does not correlate well with the overall fMRI signal strength, which suggests that the observed radial organization is not likely to result from macroscopic vasculature differences. Nevertheless, this lack of correlation does not exclude the possibility that more subtle vasculature differences within the hIC (e.g. between its central and lateral portions) may underlie the observed tuning width patterns. Furthermore, it should



be considered that we cannot distinguish between the three nuclei comprising the IC. Consequently, the broadly tuned IC regions could also reflect the dorsal or external portion of the IC. Keeping these intrinsic caveats in mind, it is nevertheless noteworthy that the interpretation of our fMRI tuning width maps in terms of neuronal population responses is compatible with reports of spectral tuning width in the cat IC, where tuning width was found to be most narrow on the medial third of frequency band laminae (Schreiner and Langner, 1988). Future studies may shed light on the possible confounds, for example by discriminating IC subdivisions based on their anatomical connectivity and/or functional properties which may be assessed non-invasively.

In most electrophysiology studies, recordings of neuronal responses are limited to single cortical or midbrain regions and consequently comparing the width of frequency tuning across multiple stages of the auditory pathway remains challenging (Bartlett et al., 2011). The large brain coverage achievable with fMRI allowed us to compare, within each single subject, the tuning width of the hIC and auditory cortical regions from responses simultaneously collected to the same natural sounds. We observed that the IC had a significantly larger proportion of voxels narrowly tuned compared to the cortex, with the lateral belt exhibiting the largest proportion of broadly tuned voxels. Voxels' tuning widths as estimated here extended over a wide distribution of values but were limited – at their lower boundary - to values broader than  $\sim 1/2$  (0.45) of an octave. These results are in apparent contradiction with recent human and monkey electrophysiology data (Bitterman et al., 2008; Bartlett et al., 2011), which showed cortical spectral tuning width as narrow as  $1/12$  octave. However, the estimation of tuning width with fMRI might differ due to several methodological factors. First, fMRI has a much poorer spatial resolution than electrophysiology, and each voxel of  $3.3 \text{ mm}^3$  of cortex contains several thousands of neurons. A voxel containing neurons with many diverse best frequencies would be assigned with a broader width than a voxel containing neurons with more similar best frequencies, which could be reflected in our results. Also, a voxel response may not represent well the heterogeneous responses of many neurons with differently shaped receptive fields, thus introducing noise and imperfection in tuning width estimation. Second, neuronal tuning width may depend on the loudness of the stimuli used. While most electrophysiological studies have measured tuning width at 10 dB or best level, here we have used much louder ( $\sim 60$  dB) complex sounds. As the tuning of a majority of neurons becomes broader with increasing sound level, a direct comparison across studies is not possible. Third, our tuning width computation reflects the width of the main spectral peak only. As additional spectral peaks in the voxels' profiles are disregarded by our fitting procedure, voxels with complex profiles would be labelled as narrowly tuned by our

method (as long as their main frequency peak is narrow). Consequently, in our measure the tuning width of a region is unrelated to its spectral complexity.

### **Contribution of ultra-high field imaging**

In our study, fMRI at 7 Tesla allowed a significant increase of imaging resolution compared to conventional measurements ( $3.3 \text{ mm}^3$  compared to the standard  $8\text{-}27 \text{ mm}^3$ ), and additionally minimized partial volume effects and contamination between locations with different frequency responses. While a  $1.5 \text{ mm}$  isotropic resolution acquisition may be possible also at lower magnetic fields (3T; Langers and van Dijk, 2011), the use of high fields presents several advantages. The (supra-linear) increase of the blood oxygen level dependent (BOLD) effect with field strength resulted in better temporal signal to noise ratio (Yacoub et al., 2001; Uludag et al., 2009), and the smaller point spread function at higher fields resulted in more spatially specific responses (Engel et al., 1997; Shmuel et al., 2007). Thus, the acquisition of functional data at ultra high fields (7 Tesla) and the experimental design with multiple types of stimuli were key factors for imaging the fine-grained functional organization of this small sub-cortical structure.

## References

- Aitkin LM, Martin RL (1987) The representation of stimulus azimuth by high best-frequency azimuth-selective neurons in the central nucleus of the inferior colliculus of the cat. *J Neurophysiol* 57: 1185-1200.
- Baumann S, Griffiths TD, Rees A, Hunter D, Sun L, Thiele A (2010) Characterisation of the BOLD response time course at different levels of the auditory pathway in non-human primates. *NeuroImage* 50: 1099-1108.
- Baumann S, Griffiths TD, Sun L, Petkov CI, Thiele A, Rees A (2011) Orthogonal representation of sound dimensions in the primate midbrain. *Nature Neurosci* 14: 423-425.
- Bartlett EL, Sadagopan S, Wang X (2011) Fine frequency tuning in monkey auditory cortex and thalamus. *J Neurophysiol* 106: 849-859.
- Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I (2008) Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451: 197-201.
- Cheung MM, Lau C, Zhou IY, Chan KC, Zhang JW, Fan SJ, Wu EX (2012) High fidelity tonotopic mapping using swept source functional magnetic resonance imaging. *NeuroImage* 61: 978-986.
- Cheung SW, Bedenbaugh PH, Nagarajan SS, Schreiner CE (2001) Functional organization of squirrel monkey primary auditory cortex: responses to pure tones. *J Neurophysiol* 85: 1732-1749.
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci* 31: 14067-14075.
- Ehret G, Merzenich MM (1988) Complex sound analysis (frequency resolution, filtering and spectral integration) by single units of the inferior colliculus of the cat. *Brain Research* 472: 139-163.
- Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex* 7: 181-192.
- FitzPatrick KA (1975) Cellular architecture and topographic organization of the inferior colliculus of the squirrel monkey. *J Comp Neurol* 164: 185-207.
- Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science* 322: 970-973.
- Formisano E, Kim D, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40: 859-869.
- Guimaraes AR, Melcher JR, Talavage TM, Baker JR, Ledden P, Rosen BR, Kiang NY, Fullerton BC, Weisskoff RM (1998) Imaging subcortical auditory activity in humans. *Human Brain Mapping* 6: 33-41.
- Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441: 197-222.
- Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394: 475-495.
- Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. *NeuroImage* 50: 1202-1211.
- Imaizumi K, Priebe NJ, Crum PA, Bedenbaugh PH, Cheung SW, Schreiner CE (2004) Modular functional organization of cat anterior auditory field. *J Neurophysiol* 92: 444-457.
- Kajikawa Y, de La Mothe L, Blumell S, Hackett TA (2005) Comparison of neuron response properties in areas A1 and CM of the marmoset monkey auditory cortex: tones and broadband noise. *J Neurophysiol* 93: 22.
- Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. *Nature* 452: 352-355.
- Kusmirek P, Rauschecker JP (2009) Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J Neurophysiol* 102: 1606-1622.
- Langers DR, van Dijk P (2011) Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cereb Cortex* 22: 2024-2038.

- Malmierca MS, Izquierdo MA, Cristaudo S, Hernandez O, Perez-Gonzalez D, Covey E, Oliver DL (2008) A Discontinuous Tonotopic Organization in the Inferior Colliculus of the Rat. *J Neurosci* 28: 4767-4776.
- Malmierca MS, Rees A, Le Beau FE, Bjaalie JG (1995) Laminar organization of frequency-defined local axons within and between the inferior colliculi of the guinea pig. *J Comp Neurol* 357: 124-144.
- Merzenich MM, Brugge J (1973) Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Research* 50: 275-296.
- Merzenich MM, Reid MD (1974) Representation of the cochlea within the inferior colliculus of the cat. *Brain Research* 77: 397-415.
- Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA (2008) Predicting human brain activity associated with the meanings of nouns. *Science* 320: 1191-1195.
- Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning and relation to voice-sensitivity. *J Neurosci* 32: 14205-14216.
- Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. *NeuroImage* 56: 400-410.
- Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. *Neuron* 63: 902-915.
- Petkov CI, Kayser C, Augath M, Logothetis NK (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biology* 4: e215. doi:10.1371/journal.pbio.0040215.
- Rauschecker JP, Tian B (2004) Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91: 2578.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111-114.
- Read HL, Winer JA, Schreiner CE (2001) Modular organization of intrinsic connections associated with spectral tuning in cat auditory cortex. *Proc Nat Acad Sci U S A* 98: 8042.
- Riecke L, van Opstal AJ, Goebel R, Formisano E (2007) Hearing illusory sounds in noise: sensory-perceptual transformations in primary auditory cortex. *J Neurosci* 27: 12684-12689.
- Schreiner CE, Langner G (1988) Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* 60: 1823-1840.
- Schreiner CE, Langner G (1997) Laminar fine structure of frequency organization in auditory midbrain. *Nature* 388: 383-386.
- Semple MN, Aitkin LM (1979) Representation of sound frequency and laterality by units in central nucleus of cat inferior colliculus. *J Neurophysiol* 42: 1626-1639.
- Shmuel A, Yacoub E, Chaimow D, Logothetis NK, Ugurbil K (2007) Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. *NeuroImage* 35: 539-552.
- Striem-Amit E, Hertz U, Amedi A (2011). Extensive Cochleotopic Mapping of Human Auditory Cortical Fields Obtained with Phase-Encoding fMRI. *PLoS ONE* 6: e17832. doi:10.1371/journal.pone.0017832.t001.
- Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, et al. (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol* 91: 1282-1296.
- Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315-2331.
- Uludag K, Muller-Bierl B, Ugurbil K (2009) An integrative model for neuronal activity-induced signal changes for gradient and spin echo functional imaging. *NeuroImage* 48: 150-165.
- van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. *Neuron* 43: 271-282.
- van De Moortele PF, Auerbach EJ, Olman C, Yacoub E, Ugurbil K, Moeller S (2009) T1 weighted brain images at 7 Tesla unbiased for Proton Density, T2 contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. *NeuroImage* 46: 432-446.

Versnel H, Zwiers MP, van Opstal AJ (2009) Spectrotemporal Response Properties of Inferior Colliculus Neurons in Alert Monkey. *J Neurosci* 29: 9725-9739.

Winer JA and Schreiner CE (2005) *The Inferior Colliculus*. New York: Springer.

Woods DL, Herron TJ, Cate AD, Yund EW, Stecker GC, Rinne T, Kang X (2010) Functional properties of human auditory cortical fields. *Front Sys Neurosci* 4: 155. doi:10.3389/fnsys.2010.00155.

Yacoub E, Shmuel A, Pfeuffer J, van de Moortele PF, Adriany G, Andersen P, Vaughan JT, Merkle H, Ugurbil K, Hu X (2001) Imaging brain function in humans at 7 Tesla. *Magn Reson Med* 45: 588-594.

### **Acknowledgements**

This work was supported in part by the National Institutes of Health (grants, R01 EB000331, P30 NS057091, (NCRR) P41 RR08079 and (NIBIB) P41 EB015894), the W.M. Keck Foundation, and MIND institute. The 7 T magnet purchase was funded in part by NSF DBI-9907842 and NIH S10 RR1395.

## **Chapter 4**

### **Processing of natural sounds: Characterization of multi-peak spectral tuning in human auditory cortex**

Moerel M, De Martino F, Santoro R, Ugurbil K, Goebel R, Yacoub E, Formisano E. Processing of natural sounds: Characterization of multi-peak spectral tuning in human auditory cortex (submitted).



## Abstract

We examine the mechanisms by which the human auditory cortex processes the frequency content of natural sounds. Through mathematical modelling of ultra-high field (7T) functional MRI responses to natural sounds, we derive frequency-tuning curves of cortical neuronal populations. With a data-driven analysis, we divide the auditory cortex into five spatially distributed clusters, each characterized by a spectral tuning profile. Beyond populations with single-peaked tuning (grouped into two clusters), approximately 60% of auditory populations are sensitive to multiple frequency bands. Specifically, we observe sensitivity to multiple frequency bands (I) at exactly one octave distance from each other, (II) at harmonically related frequency intervals, and (III) with no apparent relationship. We propose that beyond the well-known cortical tonotopic organization, multi-peaked spectral tuning amplifies selected combinations of frequency bands. Such selective amplification might serve to detect complex sound features, aid in segregating auditory scenes, and explain prominent perceptual phenomena such as octave invariance.



## Introduction

The sounds and auditory scenes we encounter every day consist of a rich and complex combination of frequencies. At the sensory periphery (cochlea) and in the sub-cortical auditory relays, sound frequency bands are selectively processed in spatially segregated channels. At the early stages of the auditory processing hierarchy, neurons with similar frequency preference cluster together and form a cochleotopic or tonotopic map (King and Nelken, 2009). This topographic organization of sound frequency is maintained at the level of the auditory cortex, where multiple tonotopic maps can be discriminated (Formisano et al., 2003; Da Costa et al., 2011; Humphries et al., 2010; Merzenich et al., 1973; Merzenich and Brugge, 1973; Moerel et al., 2012).

How frequency information is represented and processed at the level of the human auditory cortex beyond this tonotopic representation remains largely unknown. Results from invasive studies of animal audition suggest that next to their preferred frequency, cortical neurons exhibit sensitivity to additional frequency bands (e.g. bat: Fitzpatrick et al., 1993; cat: Sutter & Schreiner, 1991; Norena et al., 2008; marmoset: Kadia & Wang, 2003; Sadagopan & Wang, 2009). This sensitivity to multiple frequency bands is shaped by the acoustic environment. For example, in the bat neurons were found that showed facilitative responses to the frequencies in the bats' pulse and echo, which this animal uses during the search and pursuit of insects (Fitzpatrick et al., 1993). In marmosets and songbirds, neurons finely tuned to informative features of conspecific vocalizations were reported (Woolley et al., 2005; Wang et al., 1995). It has been hypothesized that these neurons with sensitivity to multiple frequency bands could serve as complex feature detectors, signalling the presence of a specific and informative combination of behaviourally relevant frequencies. Consequently, such neuronal tuning may play a crucial role in the creation of a more abstract, higher-level sound representation (Wang, 2007; deCharms et al., 1998; Sadagopan & Wang, 2009).

Is a similar mechanism of complex spectral tuning in place in the human auditory cortex? In this study, we use ultra-high field functional MRI (7T) to measure brain responses to a large set of natural sounds, and extract spectral profiles throughout the human auditory cortex. Next, we use a data-driven algorithm to cluster these spectral profiles and extract five spatially distributed functional subdivisions, each characterized by a specific spectral tuning profile. Simple single-peaked spectral profiles are complemented by profiles with sensitivity to multiple frequency bands. Specifically, we observe selectivity to frequency bands (I) at exactly one octave lag, (II) at harmonically related frequency intervals, and (III) with no apparent relationship to each other. We pro-

pose that this multi-peaked spectral tuning may amplify selected combinations of frequency bands, possibly serving to detect behaviourally relevant sound features, aiding in segregating auditory scenes and underlying prominent perceptual phenomena such as octave perception.

## Methods

### Subjects

Five subjects (median age = 32, 3 males) participated in the main study, and additionally took part in a separate localizer experiment (see below). The subjects had no history of hearing disorder or neurological disease, and gave informed consent before commencement of the measurements. The Institutional Review Board for human subject research at the University of Minnesota granted approval for the study.

### Stimuli

In the main study, the stimuli consisted of recordings of 168 natural sounds, and included human vocal sounds (both speech and non-speech, e.g. baby cry, laughter, coughing), animal cries (e.g. monkey, lion, horse), tool sounds and musical instruments (e.g. keys, scissors, piano, flute), and scenes from nature (e.g. rain, wind, thunder). Sounds were sampled at 16 kHz and their duration was cut at 1000 ms. In addition to the main study, we collected localizer data in the same subjects. In the localizer, the stimuli consisted of sounds grouped into eight conditions (3 tones and 5 semantic category conditions). We analyzed only the responses to the tones. Amplitude modulated tones were created in Matlab (8 Hz, modulation depth of 1) with a carrier frequency of 0.45, 0.5, and 0.55 kHz for the low frequency condition; 1.35, 1.5, and 1.65 kHz for the middle frequency condition and 2.25, 2.5, and 2.75 kHz for the high frequency conditions. Sounds were sampled at 16 kHz and their duration was cut at 800 ms.

In both the main study and the localizer, sound onset and offset were ramped with a 10 ms linear slope and their energy (RMS) levels were equalized. Inside the scanner, sounds were presented with an MR compatible audio system based on air tubes (Avotec Inc.; linear frequency transfer function up to about 4 kHz) at approximately 60 dB. Before starting the measurement, sounds were played to the subject and individual sound intensity was further adjusted in order to equalize their perceived loudness.

## Magnetic Resonance Imaging

Data were acquired on a 7T whole body system driven by a Siemens console using a head gradient insert operating at up to 80 mT/m with a slew rate of 333 T/m/s. A head RF coil (single transmit, 16 receive channels) was used to acquire anatomical ( $T_1$  weighted) and functional ( $T_2^*$  weighted BOLD) images.  $T_1$  weighted ( $1\text{ mm}^3$ ) images were acquired using a modified Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) sequence (TR=2500 ms; TI=1500 ms; TE=3.67 ms). Proton density (PD) images were acquired together with the  $T_1$  weighted images using an interleaved acquisition. The PD images were used to minimize inhomogeneities in  $T_1$  weighted images (van de Moortele, Neuroimage, 2009). Acquisition time for anatomy was around 7 minutes.  $T_2^*$  weighted functional data were acquired using an Echo Planar Imaging (EPI) sequence in which time gaps were placed after the acquisition of each volume.

The main experiment was designed according to a fast event-related scheme. The acquisition parameters were: TR = 2600 ms; time of acquisition [TA] = 1200 ms; TE = 30 ms; number of slices = 31; GRAPPA acceleration X3; partial Fourier 6/8; voxel size =  $1.5 \times 1.5 \times 1.5\text{ mm}^3$ , silent gap = 1400 ms. Slices covered the brain transversally from the inferior portion of the anterior temporal pole to the superior portion of the superior temporal gyrus bilaterally. The experiment consisted of eight runs. Six of these runs were 'train' runs, in which 144 sounds were presented 3 times in total (i.e. each sound was presented in half of the 'train' runs). The two remaining runs were 'test' runs, in which 24 different sounds were presented 6 times in total (3 times per run). Within each run, sounds were randomly spaced at a jittered interstimulus interval of 2, 3, or 4 TRs and presented – with additional random jitter - in the silent gap between acquisitions. Zero trials (trials where no sound was presented, 6% or 5% of the trials in 'train' and 'test' runs respectively), and catch trials (trials in which the sound which was just heard- was presented, 6% and 3% of the trials in 'train' and 'test' runs respectively) were included. Subjects were instructed to perform a one-back task, and were required to respond with a button press when a sound was repeated. Catch trials were excluded from the analysis. Each run lasted approximately 10 minutes.

The localizer was designed according to a blocked scheme. The acquisition parameters were: TR = 3000 ms; time of acquisition [TA] = 1500 ms; TE = 30 ms; number of slices = 44; GRAPPA acceleration X3; partial Fourier 6/8; voxel size =  $1.5 \times 1.5 \times 1.5\text{ mm}^3$ , silent gap = 1500 ms. Slices covered the brain transversally from the inferior portion of the anterior temporal pole to the superior portion of the superior temporal gyrus bilaterally. The localizer consisted of six runs, and in each run two blocks of each condition were presented. In each

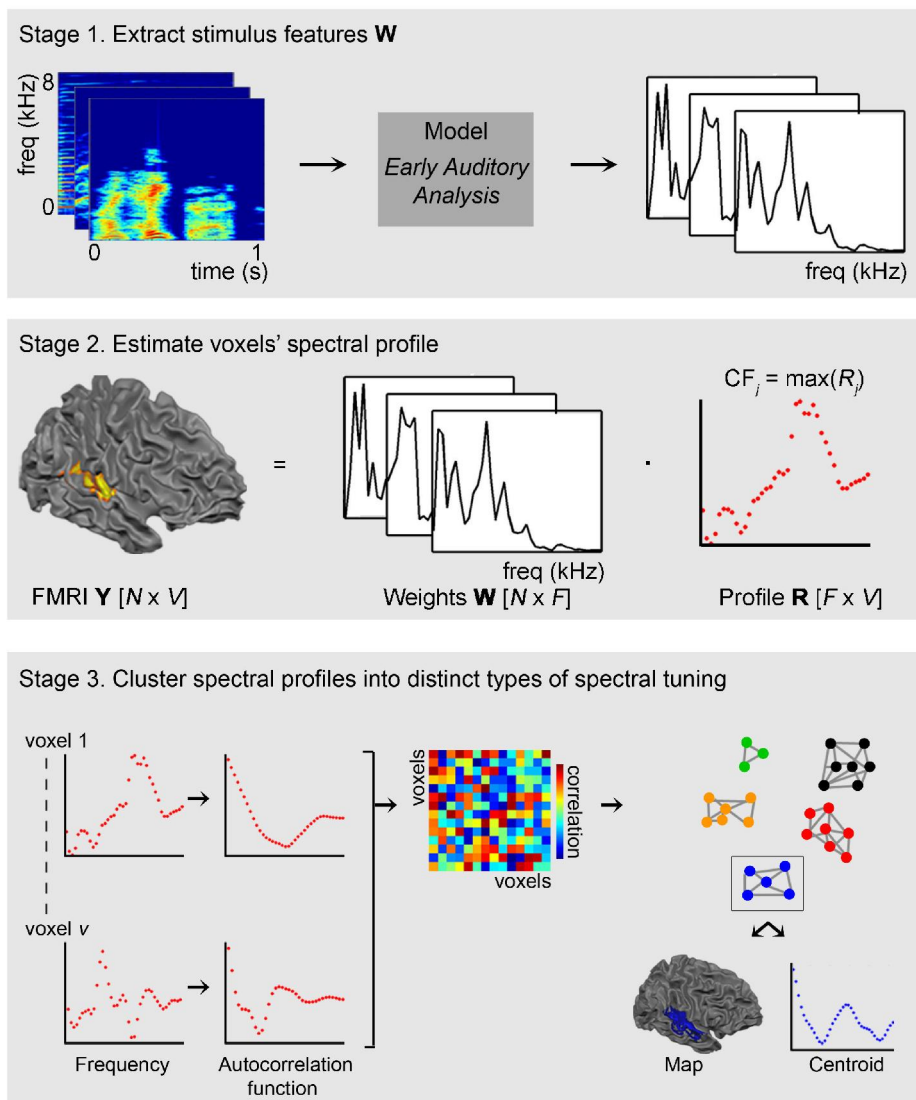
block, lasting 18 s, six sounds of the same condition were presented (one sound per TR, presented in the silent gap). Blocks of acoustic stimulation were separated from each other by 12 s of silence. Each run lasted approximately 9 minutes.

Functional and anatomical data were analyzed with BrainVoyager QX. Pre-processing consisted of slice scan-time correction (with sinc interpolation), temporal high pass filtering (removing drifts of 7 cycles or less per run for the main study, and 3 cycles or less per run for the localizer) and 3-dimensional motion correction. Additional spatial smoothing (Gaussian kernel with FWHM = 2 mm) and temporal smoothing (2 consecutive data points) were applied to the localizer data only. Functional data were co-registered to the anatomical data and normalized in Talairach space (Talairach and Tournoux, 1988). Functional data were resampled (with sinc interpolation) in the normalized space at a resolution of 1 mm isotropic. Anatomical volumes were also used to derive gray matter segmentations indicating the border between white and gray matter. Using this border, inflated hemispheres of the individual subjects were obtained.

### Computation of a voxels' spectral response profile

Based on the responses to natural sounds measured in the main study, we calculated the voxels spectral profile using customized Matlab code ([www.mathworks.com](http://www.mathworks.com)). We followed methodological procedures similar to the ones previously described for the analyses of visual responses to natural scenes (Kay et al., 2008b; Naselaris et al., 2009; Naselaris et al., 2010), which we have recently adapted and validated for the analysis of natural sounds (Morel et al., 2012; De Martino et al., 2013; for details see chapter 2).

In this analysis, sounds used as stimuli in our experiment are filtered through a biologically plausible computational model of auditory processing from the cochlea to the midbrain (NSL Tools package, available at <http://www.isr.umd.edu/CAAR/pubs.html>; Chi et al., 2005; 'Stage 1' in Figure 1). This results in the mathematical representation of the training sounds in terms of an  $N \times F$  matrix  $\mathbf{W}$  of coefficients, where  $N$  = number of sounds, and  $F$  = the number of resulting frequency bins ( $N = 144$  and  $F = 128$ ; centre frequencies ranging from 0.2 to 7 kHz). Based on the training data only, and repeated using five cross-validations (24 sounds were left out per cross-validation), we calculated the voxels spectral preference (matrix  $\mathbf{R}$  [ $F \times V$ ], where  $V$  = number of voxels) using the fMRI response matrix  $\mathbf{Y}$  [ $N \times V$ ] (Kay et al., 2008a) and the frequency representation of the sounds  $\mathbf{W}$  [ $N \times F$ ]. For each voxel  $j$ , its frequen-



**Figure 1. Encoding approach to estimate spectral response profiles**

First, sound features were extracted as the output of a computational model of early auditory processing (Chi et al., 2005). Second, measured brain responses  $\mathbf{Y}$  to the sounds' spectral components  $\mathbf{W}$  allowed estimating – through regularized regression – spectral profiles  $\mathbf{R}$  of all cortical locations ( $N$  = number of sounds;  $F$  = number of frequencies;  $V$  = number of locations/voxels). Third, we divided the voxels into clusters. We computed the autocorrelation function of each voxels' profile, and correlated voxels to represent their similarity. The data-driven Louvain clustering algorithm provided a separation of this network into clusters, each consisting of a centroid and a map.

cy profile  $R_j [(F \times 1)]$  was obtained as the relevance vector machine (RVM) solution to the linear problem:

$$Y_j = \mathbf{W} \cdot R_j$$

where each element  $i$  of the vector  $R_j$  describes the contribution of the frequency bin  $i$  to the overall response of voxel  $j$  ('Stage 2' in Figure 1; for details see Moerel et al., 2012).

### Sound prediction

From the main study, we obtained an estimate of the voxels' spectral profiles for each of the five cross-validations. We evaluated these profiles by averaging across cross-validations, and using resulting profiles to predict the voxels' response to novel sounds (i.e. sounds presented in the test runs). That is, we the spectral profiles, to predict the response  $\hat{Y}_j$  to the 24 test sounds as

$$\hat{Y}_j = \mathbf{W} \cdot R_j$$

where  $\mathbf{W} [N \times F]$  is the frequency representation of the test sounds ( $N = 24$ ). We evaluated prediction accuracy by concatenating  $\hat{Y}_j$  across voxels resulting in predicted responses  $\hat{\mathbf{Y}} [N \times V]$ , and computing correlation matrix  $\mathbf{C} [N \times N]$  between predicted and measured responses to the test sounds. Next, for each sound  $i$  we obtained the rank  $r_i$  of  $C_{i,i}$ . Note that a rank of 1 indicates perfect prediction, while a rank of  $N$  would be the worst outcome. Prediction accuracy  $P_i$  of each sound  $i$  was defined as

$$P_i = 1 - \left( \frac{r_i - 1}{N - 1} \right)$$

Values of  $P_i$  range between 0 and 1, with 1 being perfect prediction. Prediction accuracy was calculated as the average across  $N$  test sounds. To test whether this accuracy was significantly greater than zero, we compared it to the null distribution obtained using permutation testing. Specifically, we permuted  $V$  in  $\mathbf{R}$  (number of permutations = 1000) and used this permuted matrix to predict the response to test sounds. A prediction accuracy for which less than 5% of the permuted accuracies were higher, was considered significant.

### Computation of tonotopic maps

Per cross-validation, we obtained a tonotopic map by considering the characteristic frequency (CF) of a voxel as the frequency corresponding to the maximum of the coefficients in  $R_j$ . These maps were averaged across cross-validations to create one map of tonotopy per subject. Tonotopic cortical maps were obtained

by logarithmic mapping of best-frequency values to a red-yellow-green-blue colour scale (see Figure 2B).

From the localizer, tonotopic maps were extracted in the following manner. Using a single subject General Linear Model (GLM) analysis with a standard hemodynamic response model (Friston et al., 1995), we computed the responses to the three centre frequencies (.5; .1.5; 2.5 kHz) in all six runs separately. Voxels that showed a significant response to the sounds were selected ( $Q[FDR] < 0.05$ ), and the response to the three tones was z-normalized across these voxels. For each voxel, its best frequency was determined in 6-fold cross-validation (one run was left out in each fold). If the estimated best frequency had a majority across folds (3 or more occurrences), the voxel was colour-coded accordingly.

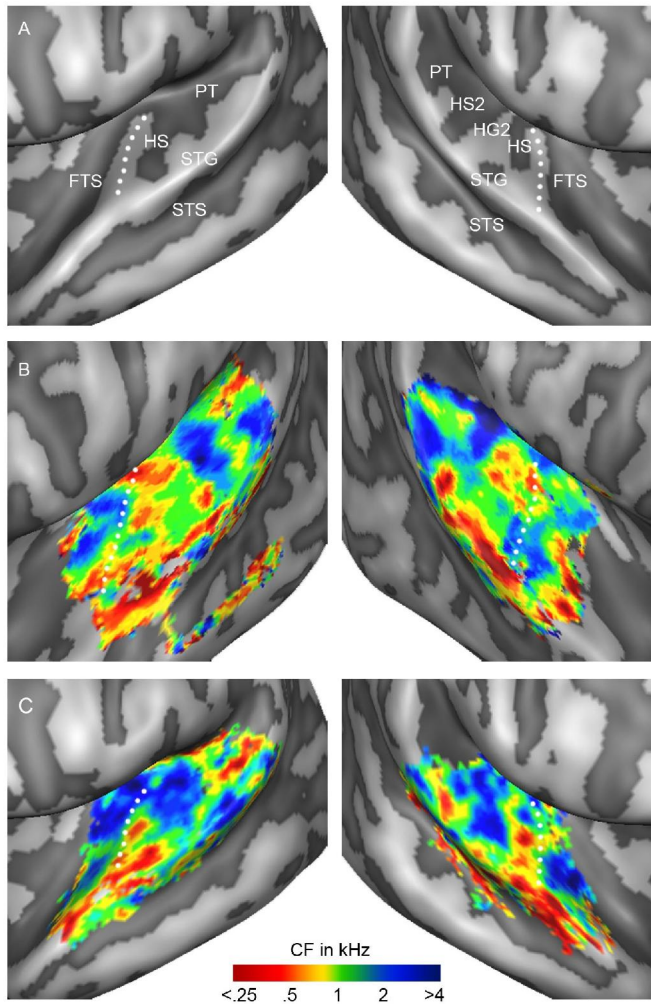
### Functional cortex based alignment

Alignment across subjects for the purpose of computing group maps is particularly challenging in the auditory cortex, due to considerable inter-individual differences in macro-anatomical landmarks and related functional responses (Da Costa et al., 2011, Penhune et al., 1996). Here, we optimized the alignment across subjects by performing functionally informed cortex based alignment (fCBA). FCBA complements standard CBA (Goebel et al., 2006; Frost and Goebel, 2011) by allowing individually defined functional regions, in addition to the major sulci and gyri, to drive the across-subject alignment.

We based fCBA on tonotopic maps obtained from the localizer. In each subject and hemisphere, we delineated the low frequency region consistently present in the vicinity of Heschl's gyrus as region of interest. FCBA was partially driven by this functional region (weighting decreased over iterations), and partially by anatomical information (weighting increased over iterations; Frost & Goebel, submitted). The resulting alignment information was used for calculating and displaying group tonotopic maps and cluster group maps (see below) as obtained from the main study.

### Extraction of spectral clusters

Maps of tonotopy reflect the main peak in voxels' spectral profiles only. In order to characterize the complex, multi-peaked spectral tuning of voxels throughout the auditory cortex (i.e. voxels' spectral tuning beyond their main frequency peak), we performed the following analysis. First, we computed the normalized autocorrelation function (autocorrelation at zero lag equal to 1) of each profile  $R_j$



**Figure 2. Tonotopic maps**

(A) Anatomy of the left and right hemisphere superior temporal plane. Light/dark colors indicate the location of gyri and sulci respectively. The abbreviations indicate the location of the first transverse sulcus (FTS), intermediate sulcus (SI), second Heschl's gyrus (HG2), Heschl's sulcus (HS), second Heschl's sulcus (HS2), planum temporale (PT), superior temporal gyrus (STG), and superior temporal sulcus (STS). White dotted lines indicate the location of HG. (B-C) Individual (subject 3) and group maps of tonotopy respectively, extracted as the maximum of the voxels' spectral profiles. The group topographic maps represent the mean across individuals aligned in fCBA space, and are shown for voxels included in  $\geq 3$  individual maps.

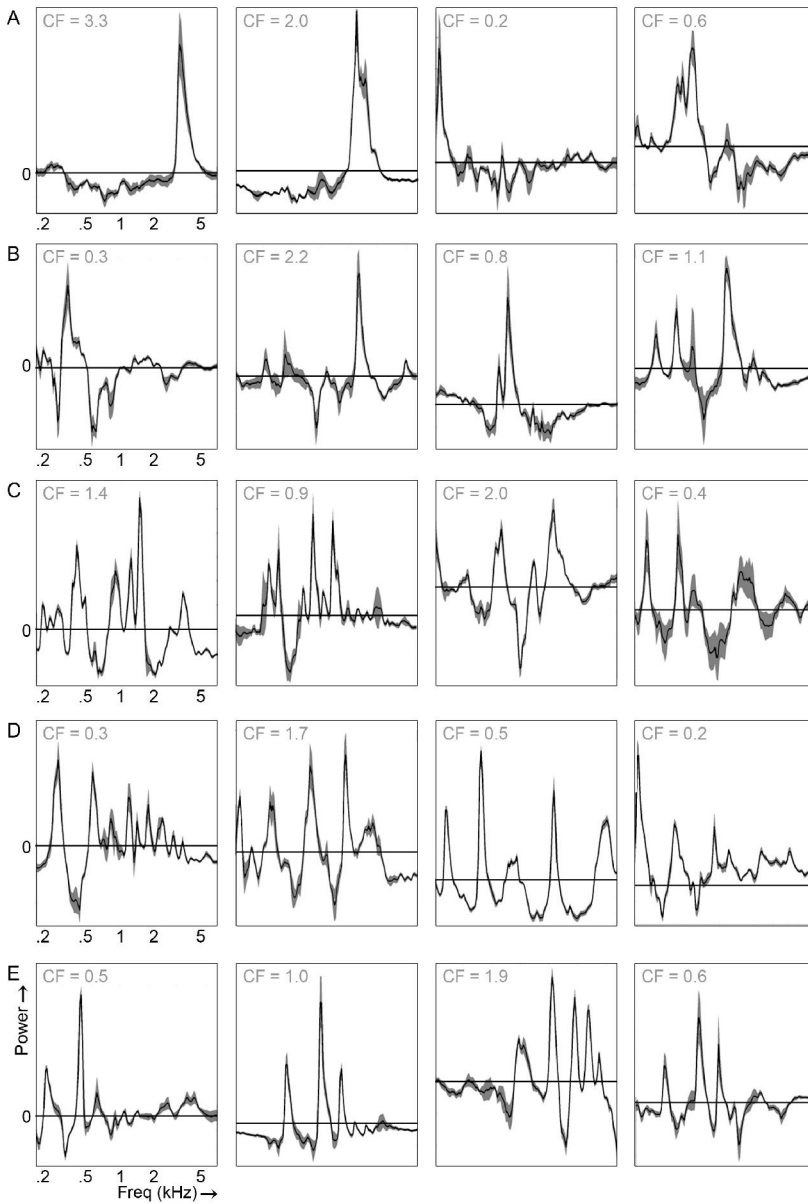
(averaged across the five cross-validations). Note that by computing the profiles' autocorrelation function, we obtained a representation of spectral profiles that reflects spectral modulations and is independent of the main frequency peak. Second, we used a data-driven clustering algorithm to obtain the different



'types' of spectral profiles present in the auditory cortex. Specifically, we computed the correlation between the autocorrelation profiles of the voxels, reflecting the voxels' similarity in spectral modulation space (sparseness of this network was set at 0.3, and the network contained only positive values). Using the Louvain algorithm as implemented in the Brain Connectivity Toolbox (<http://www.brain-connectivity-toolbox.net>, an open-access Matlab network analysis toolbox), we obtained a fully data-driven subdivision of this network into spatially non-overlapping clusters (see 'Stage 3' in Figure 1; Blondel et al., 2008; Rubinov and Sporns, 2011).

As the output of this algorithm depends on heuristics, it finds a slightly different solution from run to run. In order to evaluate the stability of the obtained solutions, we ran the algorithm multiple times ( $N = 16$  repetitions). Additionally, we evaluated the output of the Louvain algorithm after randomizing the correlation matrix (by randomizing the phase of the Fourier transformed matrix, and computing the inverse Fourier transform;  $N = 16$  repetitions; Zalesky et al., 2012). We evaluated the output of the Louvain algorithm by the modularity of the obtained subdivision, its stability in number of resulting clusters across runs, and its stability across runs (measured as estimated mutual information across repetitions), both when computed on the data and on the randomized network (Rubinov and Sporns, 2011).

After this quality check on the output of the Louvain algorithm, we computed the characteristic profile (i.e. the centroid) of each resulting cluster as follows. We normalized each voxels' spectral profile by expressing all amplitude and frequency values as ratios with respect to its amplitude maximum and the frequency at that maximum. Thus, the abscissa and ordinate became, respectively,  $F_n = F/F_{\max}$  and  $A_n = A/A_{\max}$ , where  $A_n$  and  $F_n$  were the normalized values (Schwartz et al., 2003).  $F_n$  ranged between  $0.1 \cdot F_{\max}$  and  $10 \cdot F_{\max}$ , divided in 134 logarithmically spaced bins. By normalizing the spectral profiles in this manner, both their maxima and additional frequency peaks below and above the maxima were aligned across voxels. Each clusters' centroid was computed by averaging the normalized spectral profiles of the voxels in that cluster. Based on the centroids, we identified 'matching' clusters across subjects (mean [SD] correlation between matching centroids =  $0.99 [9.7 \cdot 10^{-3}]$ ). Group centroids were obtained by averaging matching centroids across subjects. Peaks in each centroid were tested for significance by performing a one-sample t-test in each bin of  $F_n$ .



**Figure 3. Representative spectral profiles**

The voxels' sensitivity to frequencies spanning 5.3 octaves (~0.2 to 7 kHz), of four representative voxels per cluster are shown (data of subject 5). Examples of the (A) broadly tuned, single peaked cluster; (B) cluster with simple tuning and pronounced inhibitory sidebands surrounding the main frequency peak; (C) Multi-peaked cluster without clear distance between frequency peaks; (D) Multi-peaked octave cluster; (E) Multi-peaked harmonic cluster.

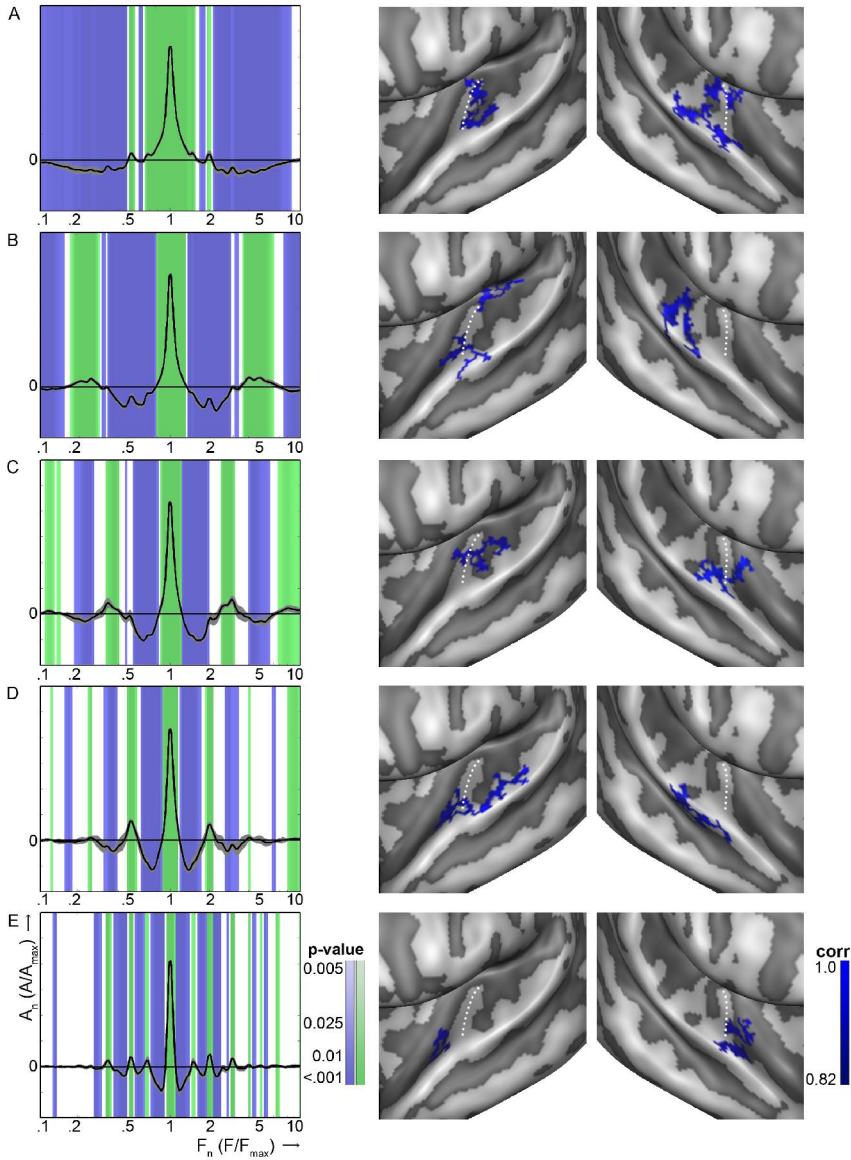
Cluster spatial maps were obtained as follows. We computed the correlation of each voxels' autocorrelation profile to the corresponding clusters' centroid (hard clustering; each voxel belongs to one cluster only), reflecting the amount to which that voxel represented the cluster. Individual subject maps were created by smoothing maps on the surface and thresholding the maps ( $r > 0.55$ ; cluster threshold  $> 10$ ). The spatial layout of individual subject maps was quantified by dividing each hemisphere in five anatomically defined sub-regions (Heschl's gyrus [HG], planum polare [PP], planum temporale [PT], rostral and caudal superior temporal gyrus [rSTG and cSTG respectively]; based on Kim et al., 2000; see Figure 6A). Next, we computed the proportion each cluster located in each of the anatomical regions. Finally, group maps were created by transforming individual maps into fCBA space, averaging matching maps across subjects (threshold on single subject maps  $r > 0.60$ ), and smoothing resulting maps on the group surface (repeat value = 1).

## Results

### Estimation of voxels' spectral profiles and extraction of tonotopic maps

As expected, sounds evoked significant activation in a large expanse of the superior temporal cortex. The activated region included early auditory areas along HG and surrounding regions on PP, PT, STG, and parts of the superior temporal sulcus (STS). Based on the estimated voxels' spectral profiles (see Methods), responses to new sounds could be predicted significantly above chance in each subject ( $p < 0.01$  in each subject; min/max/chance prediction accuracy = 0.65/0.76/0.50), illustrating that these profiles accurately characterize the voxels' frequency tuning.

Based on the maximum of each voxels' spectral tuning curve, tonotopic maps were consistently extracted across subjects and hemispheres (maps of a single subject and group maps are displayed in Figure 2). Consistent with results of previous studies, we observed a large low frequency region near HG, bordered anteriorly (on HG and first transverse sulcus [FTS]) and posteriorly (on Heschl's sulcus [HS] and anterior PT) by regions preferring higher frequencies (Da Costa et al., 2011; Formisano et al., 2003; Humphries et al., 2010; Moerel et al., 2012). Beyond Heschl's region, on PP, PT, and STG/STS, additional frequency clusters were present (see Figure 2). Tonotopic maps reflect only the main peak in the voxels' spectral profiles, and thereby represent well those voxels that displayed a simple, single peaked frequency profile (for example



**Figure 4. Group cluster maps**

Centroids (left) and group maps (right) characterizing the five extracted clusters (Rows A-E consecutively display the broad cluster, cluster with attenuation bands, multi-peaked cluster with no clear relationship between peaks, octave cluster, and harmonic cluster). The centroids' main peak represents the main frequency peak across voxels, and additional peaks show the presence of sensitivity to additional frequency bands at consistent spectral intervals. Positive and negative significant deviations from zero are colour coded in green and blue respectively. The maps show the correlation of each voxels' profile to the clusters' centroid, averaged across subjects. Maps are displayed on an fCBA-based reconstruction of the temporal cortex, and white dotted lines indicate the location of HG.

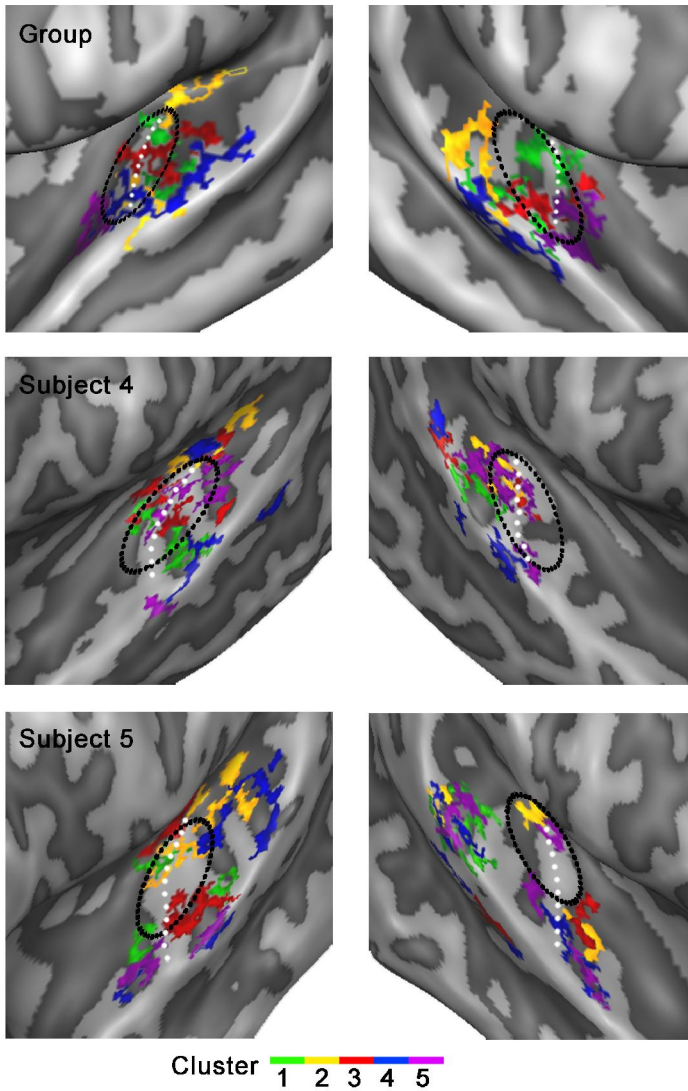
Figure 3A). However, most voxels had a more complex multi-peaked frequency preference (Figure 3C-E). Multi-peaked voxels were present both on HG (e.g. Figure 3C, 2nd column) and beyond HG, on PT and STG/STS.

### Cluster extraction with Louvain algorithm

In order to examine the spectral profiles beyond their main tonotopic peak, we used a data-driven algorithm to divide the voxels into separate clusters according to their type of spectral tuning (see Stage 3 in Figure 1; Blondel et al., 2008; Rubinov and Sporns, 2011). The algorithm's solution had a high modularity in each subject (mean[SE] =  $0.25[2.3 \cdot 10^{-3}]$ ), a stable number of extracted clusters across repeated runs of the algorithm (median[SE] =  $5[0.01]$ ), and high estimated mutual information across resulting partitions (mean[SE] =  $0.95[0.05]$ ). When running the same algorithm on a randomized network, modularity decreased (mean[SE] =  $0.15[0.01]$ ), variability in number of extracted clusters increased (median[SE] =  $3[3.71]$ ), and estimated mutual information across resulting partitions decreased (mean[SE] =  $0.19[0.01]$ ), ensuring that the clustering algorithm provided a satisfactory partitioning.

### Cluster centroids: Characteristic tuning profiles

The algorithm divided auditory cortical voxels into five clusters, based on their spectral tuning. Each of the five clusters contained respectively 19.9, 17.8, 19.0, 16.7, and 26.6 percent of total number of voxels (average across subjects). Resulting characteristic profiles (i.e. centroids) of each resulting cluster were highly consistent across subjects (mean [SD] correlation between corresponding centroids =  $0.99 [9.7 \cdot 10^{-3}]$ ). The first centroid (Figure 4A, left column) included voxels with broadly tuned frequency profiles (e.g. voxels in Figure 3A). The second centroid (Figure 4B, left column) described a cluster with significant positive peaks around  $0.25 \cdot F_{\max}$  and  $4 \cdot F_{\max}$ , corresponding to a distance of two octaves (i.e., three harmonics) between the maximum frequency and additional peaks. Voxels included in this cluster often showed one distinct frequency peak flanked by pronounced attenuation sidebands (e.g. voxels in Figure 3B). The third centroid (Figure 4C, left column) displayed significant positive frequency peaks around  $0.35 \cdot F_{\max}$  and  $3 \cdot F_{\max}$ , and additionally around  $0.1 \cdot F_{\max}$  and  $9 \cdot F_{\max}$ . Thus, this centroid displayed sensitivity to additional frequency bands at a distance of two harmonics from the maximum frequency. Voxels in this cluster often had multi-peaked spectral profiles, with no clear relationship between the peaks (e.g. voxels in Figure 3C). The fourth centroid (Figure 4D, left column)



**Figure 5. Group and individual cluster maps**

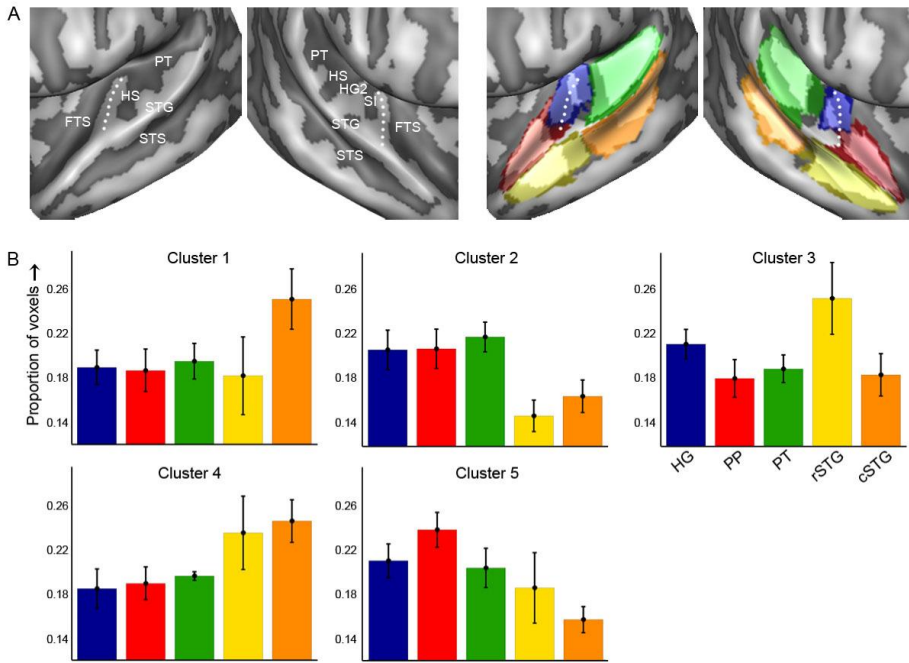
Clusters as extracted from the group, for subject 4 and for subject 5 are shown in the top, middle and bottom row respectively. The colours show maps belonging to clusters 1-5 in green, yellow, red, blue and purple respectively (i.e. octave and harmonic cluster shown in blue and purple respectively). Maps are smoothed on the surface (repeat value = 1). Consequently maps are thresholded based on the voxels' correlation to the centroid ( $> 0.82/ > 0.55$  for group and individual subject maps respectively) and cluster thresholded (minimum threshold = 10 voxels). The maps are displayed on a reconstruction of the temporal cortex at individual subject level. White dotted lines indicate the location of HG. Black dotted lines indicate the location of the PAC, identified as the main high-low-high tonotopic gradient in the proximity of HG in individual tonotopic maps.

had significant positive peaks located at  $0.5 \cdot F_{\max}$ ,  $2 \cdot F_{\max}$ , and additionally at  $0.25 \cdot F_{\max}$  and  $4 \cdot F_{\max}$ . Voxels in this cluster were selective to frequency bands at octave interval (e.g. voxels in Figure 3D). Thus, the voxels belonging to this fourth cluster were not sensitive to all harmonic distances, but instead displayed a specific selectivity to octave frequency lags. We refer to this cluster as the 'octave cluster'. Finally, the fifth centroid (Figure 4E, left column) displayed a large number of localized significant positive peaks, at 0.35, 0.5, 0.65, 1.5, 2, 3, 4, 5, 7 and  $10 \cdot F_{\max}$ . Interestingly, voxels belonging to this cluster were sensitive to additional frequency bands at multiple harmonics of the lowest frequency peak. For example, the voxels in the first and second column of Figure 3E display sensitivity to 0.2/0.4/0.6 kHz and to 0.5/1.0/1.5 kHz respectively. We refer to this cluster as the 'harmonic cluster'.

### Cluster spatial maps

After investigating the five centroids, we explored the cortical location of each cluster at group level and in individual subjects (see right column of Figure 4, Figure 5, and Figure 6). At group level, we observed that the first cluster (Figure 4A, green in Figure 5) occupied posterior primary auditory cortex (PAC; black dotted circles in Figure 5) and cSTG. The second cluster (Figure 4B, yellow in Figure 5) occupied parts of medial HG, medial HS, and PT. The third cluster (Figure 4C, red in Figure 5) was located on HG (occupying the lateral part of PAC). The octave cluster (Figure 4D, blue in Figure 5) was located lateral to PAC, occupying the STG. Finally, the harmonic cluster was located antero-lateral to PAC (Figure 4E, purple in Figure 5), and occupied a small region on PP.

The spatial distribution of the clusters in individual subjects was quite variable (Figure 5, second and third row), reflected in the histograms of Figure 6 by the spread of voxels in each cluster across anatomical regions. However, the main peaks in these histograms followed the pattern observed in the group maps. That is, the first cluster occupied the posterior locations PT and cSTG. The second cluster was localized in medial locations, occupying regions in PP, HG, and PT. The only exception to the group pattern was seen for the third cluster (red in Figure 5), which showed a peak in the histograms within rSTG additional to the peak in HG. The octave and harmonic cluster occupied the STG (both rostral and caudal) and PP respectively, in accordance with the patterns observed in the group maps (compare top rows of Figures 5 and 6).



**Figure 6. Spatial distribution of clusters**

(A) Anatomy of the left and right hemisphere superior temporal plane (left) and subdivision into five regions (based on Kim et al., 2000; right). HG, PP, PT, rostral and caudal STG (rSTG and cSTG) are indicated in blue, red, green, yellow, and orange respectively. Overlap in defined regions across three to five subjects is shown in a range of dark to light hues. (B) For each cluster, the proportion of voxels in each of the anatomical regions is shown. Clusters 1-5 consecutively display the broad cluster, cluster with attenuation bands, multi-peaked cluster with no clear relationship between peaks, octave cluster, and harmonic cluster. Black bars indicate standard error across subjects.

## Discussion

In this study, we used ultra-high field functional MRI to extract neuronal populations' spectral profiles based on their responses to natural sounds. With a data-driven analysis, we identified five clusters of spectral tuning types. Two of these clusters displayed simple single-peaked spectral profiles, containing a broadly or narrowly tuned frequency peak. However, approximately 60% of neuronal populations throughout auditory cortex (divided between the three remaining clusters) displayed complex frequency selectivity beyond their main tonotopic frequency peak. We identified neuronal populations with sensitivity to multiple frequency bands (I) at exactly one octave distance from each other, (II) at harmonically related frequency intervals, and (III) with no apparent relationship. We



propose that beyond the well-known tonotopic organization of the auditory cortex, this multi-peaked spectral tuning concurs to define the representation space of natural sounds in the human auditory cortex.

### **Encoding to characterize spectral tuning in human auditory cortex**

We used an fMRI ‘encoding’ procedure and responses to natural sounds (Kay et al., 2008b; Moerel et al., 2012; Naselaris et al., 2009; Naselaris et al., 2010) to calculate the voxels’ frequency profiles, which were subsequently used to extract clusters and centroids. Using natural sounds for feature mapping (i.e. estimating the voxels’ response profile) has several advantages over the use of artificial sounds. First, as natural sounds inherently engage auditory cortical neurons in meaningful and behaviourally relevant processing, they may be optimal for studying the functional architecture of higher order auditory areas. Second, previous studies showed that response profiles estimated with artificial sounds do not predict responses to natural sounds well (Bitterman et al., 2008; Machens et al., 2004). Consequently, the use of natural sounds was advised in mapping response features (Theunissen et al., 2000). Third, the method implemented here allows the simultaneous estimation of the voxels’ feature preference (e.g. tuning to multiple frequency peaks). As neuronal populations show nonlinearities in their feature tuning, such simultaneous estimation is of paramount importance.

By mapping the main peak of the voxels’ spectral profiles, we showed frequency selective regions throughout the superior temporal cortex. In agreement with previous fMRI studies, we observed a large low frequency area in the vicinity of HG, surrounded anteriorly and posteriorly by regions preferring higher frequencies (Da Costa et al., 2011; Formisano et al., 2003; Talavage et al., 2004). Although the exact relationship between these main frequency gradients and auditory fields remains debated (Humphries et al. 2010; Langers & van Dijk, 2012), these regions most likely reflect the location of the human primary auditory cortex (i.e. the homologues of monkey primary fields AI and R; Kosaki et al., 1997; Hackett et al., 1998). Beyond these main frequency gradients, we observed frequency selective clusters in regions that – based on their underlying anatomy – probably reflect belt and parabelt cortex (PP, PT, and STG/STS). Tonotopic maps in these regions are less frequently reported, yet these results replicate findings of at least two previous studies (Moerel et al., 2012; Striem-Amit et al., 2011).

### **Types of multi-peaked spectral tuning**

In a data-driven manner, we identified five clusters of spectral tuning types. Voxels within two of these clusters displayed an overall simple tuning profile, including voxels with a broad frequency peak, and more narrowly tuned voxels with attenuating sidebands. This type of spectral tuning may reflect the traditionally reported neuronal populations with single peaks. Voxels belonging to these clusters were found throughout the superior temporal plane. Functionally, this spectral tuning might be relevant for capturing the overall frequency content of incoming sounds.

Beyond these simple spectral tuning types, we observed three clusters (approximately 60% of neuronal populations throughout auditory cortex) that displayed complex frequency selectivity beyond their main tonotopic frequency peak. Specifically, we observed a cluster in which voxels displayed selectivity to multiple octave frequency lags. Based on their spectral profiles, we predict that these octave-tuned voxels - clustered most densely along the STG - respond in similar manner to tones with frequencies at a 2:1 ratio. Consequently, they could represent the biological basis for the percept of octave generalization. A hard-wired octave representation in the brain is in accordance with the widespread occurrence, early onset, and generalization beyond the human species of octave perception (Demany and Armand, 1984; Randel, 2003; Wright et al., 2000). Moreover, our findings of octave-based tuning are in accordance with invasive recordings in various mammals, which observed octave tuning at neuronal level (Brosch et al., 1999; Brosch & Schreiner, 2010; Kadia & Wang, 2003; Norena et al., 2008).

### **Functional relevance of neuronal sensitivity to harmonic structure**

Next, we observed a complex cluster with spectral tuning that finely followed sensitivity to multiple harmonic lags. As the natural sounds most important to humans (i.e. vocalizations, music) are harmonically structured (Ross et al., 2007; Norena et al., 2008), these multi-peaked spectral profiles may be functionally relevant for amplifying the spectral content of those sounds specifically. In everyday life, we are constantly exposed to sounds from different sources, whose spectral components overlap with each other. In order to perceive sounds we must segregate them into separate auditory objects, a task which is referred to as 'auditory scene analysis' (Bregman, 1990). We accomplish this task by grouping components of the same source together based on various fundamental features of complex sounds, such as harmonicity (Darwin, 1997). For example, a stack of harmonically related frequencies is perceptually fused

into a single tone (Borchert et al., 2011; Micheyl, 2010). The multi-peaked cortical neuronal populations could underlie the perceptual mechanism of fusing harmonically related components. That is, the harmonic tuning within this spectral cluster could group sound components for the purpose of object segregation, or significantly enhance harmonic components relative to background noise without harmonic structure (Kadia and Wang, 2003).

Finally, voxels in the third multi-peaked cluster, most densely clustered on the middle region on the PAC, displayed additional sensitivity to multiple frequency bands with no clear relationship to each other. The clusters' centroid displayed sensitivity to frequency bands at 1.6 octaves from the main peak, consistent with the main peaks' second harmonic. At present, we cannot identify an acoustic or behavioural correlate of the multiple frequency bands to which the voxels in this cluster are sensitive.

### **Neuronal substrate underlying complex population tuning**

Using fMRI, we observed spectral tuning at the level of neuronal populations. Various types of spectral tuning at the level of individual neurons could underlie our observations. At present, we cannot draw conclusions regarding the nature of the sensitivity to multiple frequency bands that we observe in the human auditory cortex. First, the observed sensitivity to multiple frequency bands may reflect true multi-peaked neuronal tuning, as has been observed in a number of studies and species (e.g. bat: Fitzpatrick et al., 1993; cat: Norena et al., 2008; monkey: Kadia & Wang, 2003). Second, the sensitivity to multiple frequency bands could result from combination-sensitive neuronal mechanisms (Sadagopan and Wang, 2009; Wang et al., 2005). As opposed to truly multi-peaked neurons, combination sensitive neurons only display sensitivity to multiple frequencies when probed with that exact frequency combination (non-linear response mechanism). Third, as each measured voxel in the current study included thousands of neurons, we may observe the complex average spectral profile of many simple, single-peaked neurons. In that case, the observed multi-peaked tuning would only emerge at a population level of neuronal responses.

To advance our understanding regarding the functional relevance of the cortical spectral sound representation, a well-controlled exploration of non-linear and combination-sensitive tuning to the multiple frequency bands within these profiles is needed. Furthermore, as previous studies show that feature tuning in auditory cortex is highly affected by changes in task, context, and attention (Atiani et al., 2009; Fritz et al., 2003), the challenge for future work is to investigate changes in spectral profiles during these manipulations. Our results can guide

future explorations of complex spectral tuning in human auditory cortex, and the methods described here provide a means for this endeavour.

## References

- Atiani S, Elhilali M, David SV, Fritz JB, Shamma SA (2009) Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron* 61: 467-480.
- Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I (2008) Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451: 197-202.
- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech* P10008.
- Borchert EM, Micheyl C, Oxenham AJ (2011) Perceptual grouping affects pitch judgements across time and frequency. *J Exp Psychol Hum Percept Perform* 37: 257-269.
- Bregman AS (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge, MA: MIT Press).
- Brosch M, Schreiner CE (2000) Sequence sensitivity of neurons in cat primary auditory cortex. *Cerebral Cortex* 10: 1155-1167.
- Brosch M, Schulz A, Scheich H (1999) Processing of sound sequences in macaque auditory cortex: Response enhancement. *J Neurophysiol* 82: 1542-1559.
- Chi T, Ru P, Shamma SA (2005) Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am* 118: 887-906.
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci* 31: 12067-14075.
- Darwin CJ (1997) Auditory grouping. *Trends in Cognitive Sciences* 1: 327-333.
- deCharms RC, Blake DT, Merzenich MM (2004) Optimizing sound features for cortical neurons. *Science* 280: 1439-1443.
- De Martino F, Moerel M, van de Moortel PF, Ugurbil K, Goebel R, Yacoub E, Formisano E (2013) Spatial organization of frequency preference and selectivity in the human inferior colliculus. *Nat Comm* 22, doi: 10.1038/ncomms2379.
- Demany L, Armand F (1984) The perceptual reality of tone chroma in early infancy. *J Acoust Soc Am* 76: 57-66.
- Fitzpatrick DC, Kanwal JS, Butman JA, Suga N (1993) Combination-sensitive neurons in the primary auditory cortex of the mustached bat. *J Neurosci* 13: 931-940.
- Formisano E, Kim D, Di Salle F, van de Moortele PF, Ugurbil K, et al. (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40: 859-869.
- Friston KJ, Frith CD, Turner R, Frackowiak RS (1995) Characterizing evoked hemodynamics with fMRI. *Neuroimage* 2: 157-165.
- Fritz JB, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience* 6: 1216-1223.
- Frost M, Goebel R (2011) Measuring structural-functional correspondence: Spatial variability of specialized brain regions after macro-anatomical alignment. *Neuroimage* 59:1369-1381.
- Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp* 27: 392-401.
- Hackett TA, Stepniewska I, Kaas JH (1998) Subdivision of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394: 475-495.
- Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. *Neuroimage* 50: 1202-1211.
- Kadia SC, Wang X (2003) Spectral integration in A1 of awake primates: Neurons with single- and multi-peaked tuning characteristics. *J Neurophysiol* 89: 1603-1622.
- Kay KN, David SV, Prenger RJ, Hansen KA, Gallant JL (2008a) Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. *Hum Brain Mapp* 29: 142-156.

- Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008b) Identifying natural images from human brain activity. *Nature* 452: 352-356.
- Kim J, Crespo-Facorro B, Andreasen NC, O'Leary DS, Zhang B, Harris G, Magnotta VA (2000) An MRI-based parcellation methods for the temporal lobe. *Neuroimage* 11: 271-288.
- King AJ, Nelken I (2009) Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nature Neuroscience* 12: 698-701.
- Kosaki H, Hashikawa T, He J, Jones EG (1997) Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. *J Comp Neurol* 386: 304-316.
- Langers DR, van Dijk P (2012) Mapping the tonotopic organization of the human auditory cortex with minimally salient acoustic stimulation. *Cerebral Cortex* 22: 2024-2038.
- Machens CK, Wehr MS, Zador AS (2004) Linearity of cortical receptive fields measured with natural sounds. *J Neurosci* 24: 1089-1100.
- Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. *Science* 296: 910-913.
- Merzenich MM, Brugge JF (1973) Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res* 50: 275-296.
- Merzenich MM, Knight PL, Roth GL (1973) Representation of cochlea within primary auditory cortex in the cat. *Brain Res* 63: 343-346.
- Micheyl C, Oxenham AJ (2010) Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings. *Hearing Research* 266: 36-51.
- Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci* 32: 14205-14216.
- Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. *Neuron* 63: 902-915.
- Naselaris T, Kay KN, Nishimoto S, Gallant JL (2010) Encoding and decoding in fMRI. *Neuroimage* 56: 400-410.
- Norena A, Gourevitch B, Pienkowski M, Shaw G, Eggermont JJ (2008) Increasing spectrotemporal sound density reveals an octave-based organization in cat primary auditory cortex. *J Neurosci* 28: 8885-8896.
- Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, et al. (2012) Reconstructing speech from human auditory cortex. *PLoS Biol* 10: e1001251. doi:10.1371/journal.pbio.1001251.
- Penhune VB, Zatorre RJ, MacDonald JD, Evans AC (1996) Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb Cortex* 6: 661-672.
- Randel DM, (ed.). *The Harvard Dictionary of Music* (2003). Cambridge Massachusetts, Harvard University Press (4th ed.).
- Ross D, Choi J, Purves D (2007) Musical intervals in speech. *Proc Natl Acad Sci U S A* 104: 9852-9857.
- Rubinov M, Sporns O (2011) Weight-conserving characterization of complex functional brain networks. *Neuroimage* 56: 2068-2079.
- Sadagopan S, Wang X (2009) Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. *J Neurosci* 29: 11192-11202.
- Schwartz DA, Howe CQ, Purves D (2003). The statistical structure of human speech sounds predicts musical universals. *J Neurosci* 23: 7160-7168.
- Striem-Amit E, Hertz U, Amedi A (2011) Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLoS One* 6: DOI:10.1371/journal.pone.0017832.
- Sutter ML, Schreiner CE (1991) Physiology and topography of neurons with multi-peaked tuning curves in cat primary auditory cortex. *J Neurophysiol* 65: 1207-1226.
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Stuttgart: G. Thieme.

- Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, et al. (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol* 91: 1282-1296.
- Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315-2331.
- van De Moortele PF, et al. (2009) T1 weighted brain images at 7 Tesla unbiased for Proton Density, T2 contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. *NeuroImage* 46: 432-446.
- Wang X (2007) A sharper view from the top. *Nature Neuroscience* 10:1509-1511.
- Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli *Nature* 435: 341-346.
- Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74: 2685-2706.
- Woolley SMN, Fremouw TE, Hsu A, Theunissen FE (2005) Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience* 8: 1371-1379.
- Wright AA, Rivera JJ, Hulse SH, Melissa S, Neiworth JJ (2000). Music perception and octave generalization in rhesus monkeys. *J Exp Psychology* 129: 291-307.
- Zalesky A, Fornito A, Bullmore E (2012) On the use of correlation as a measure of network connectivity. *NeuroImage* 60: 2096-2106.

## Acknowledgements

We thank A. Goulas and M. Frost for comments and discussions. This work was supported by Maastricht University, the Netherlands Organization for Scientific Research (NWO; Top talent grant 021-002-102 (M.M.) and Innovational Research Incentives Scheme Vidi grant 452-04-330 (E.F.), the National Institutes of Health (NIH grants P41 EB015894, P30 NS076408, and S10 RR26783), and the WM KECK Foundation.

## **Chapter 5**

### **The perceptual constancy of musical notes across octaves emerges from multi-peak spectral tuning in primary auditory cortex**

Moerel M, De Martino F, Santoro R, Ugurbil K, Yacoub E, Formisano E. The perceptual constancy of musical notes across octaves emerges from multi-peak spectral tuning in primary auditory cortex (in preparation).





## Abstract

Octave generalization refers to the perceived constancy of a tone, or a musical note, with a doubling of its frequency. This is a well-known perceptual phenomenon, which lays at the foundation of melody recognition and music perception. To date, the neural underpinnings of octave generalization remain largely unknown. Here, we show that octave generalization emerges from the response patterns of finely tuned neuronal populations in human primary auditory cortex (PAC). Using 7 Tesla functional MRI, we characterize auditory cortical locations according to their spectral tuning and identify – in a data-driven manner – the locations tuned to multiple frequency bands at exactly one octave distance. With independent measurements, we then show that the brain response patterns in these octave-tuned locations – but not in other auditory cortical locations – are significantly similar for musical notes at different octave intervals. These findings show that human auditory cortex entails octave-invariant representations of musical notes. Furthermore, they indicate that neuronal representations in PAC encode the perceptual interpretation of sounds, beyond the isomorphic mapping of acoustic frequencies from the sensory environment.

## Introduction

Central to melody recognition and music perception is the phenomenon of octave generalization, which describes the perceived similarity of tones with fundamental frequency (F0) at a 2:1 ratio. The fundamental importance of the octave interval in perception is illustrated by the observation that it is the only interval common to nearly all musical scales ever evolved (Randel, 2003). Infants already generalize a melody across octaves (Devany and Armand, 1984) and even monkeys assign a special status to octaves, judging melodies transposed by multiples of octaves as being more similar to the original than non-integer transpositions (Wright et al., 2000). The widespread occurrence, early onset, and generalization beyond the human species of octave perception suggest that an octave representation may be hard-wired in the brain, yet to date its neuronal underpinnings remain elusive.

Throughout the auditory system, neurons are described by their characteristic frequency (CF), which is the frequency to which they respond best. Auditory neurons are spatially ordered according to their CF, resulting in a cochleotopic or tonotopic map. This topographic organization of frequency is maintained in the various auditory relays (King and Nelken, 2009; Merzenich et al., 1973; Merzenich and Brugge, 1973). Beyond the tonotopic organization, recent studies in various non-human mammals suggest that a subset of neurons in the primary auditory cortex may exhibit octave-based tuning properties. Recordings from cortical neurons in the primary auditory cortex of the macaque (Brosch et al., 1999) and the cat (Brosch and Schreiner, 2000) identified neurons that responded strongest when they were exposed to two sequentially-presented tone pips with a spectral difference of one octave. Besides this temporal octave tuning, cortical neurons that display sensitivity to multiple frequency bands have been shown. Interestingly, of all possible frequency distances between the multiple spectral peaks to which these neurons might be tuned, frequency distances of one octave were highly overrepresented (Kadia and Wang, 2003; Norena et al., 2008).

Functional imaging studies have shown that the human auditory cortex is tonotopically organized (Formisano et al., 2003; Da Costa et al., 2011; Humphries et al., 2010; Moerel et al., 2012), similarly to non-human mammals. Furthermore, our recent fMRI findings suggest the existence of neuronal populations (voxels) within human auditory cortex that are finely tuned to multiple frequency bands at one octave lag (Moerel et al., submitted). In the current study, we test the hypothesis that these octave-tuned neuronal populations represent musical notes in octave-invariant manner and thereby underlie the human octave percept. We use ultra-high field functional MRI (7T) to measure the brains'

responses to a large set of natural sounds, and extract the spectral tuning properties of each cortical location. In a data-driven manner, we identify those locations that display sensitivity to multiple frequency bands at one octave distance from each other (i.e. octave-tuned locations). Next, we measure the response throughout the cortex to musical notes spanning three octaves. We show that in octave-tuned populations only – which are clustered most densely on primary auditory cortex (PAC) - responses to notes one octave apart are more similar than responses to notes at other intervals, reflecting an octave-invariant representation. We propose that the universal phenomenon of octave perception is a consequence of these selectively octave-tuned neuronal populations in human auditory cortex.

## Materials and Methods

### Subjects

Six subjects participated in this study. The subjects had no history of hearing disorder or neurological disease, and gave informed consent before commencement of the measurements. The Institutional Review Board for human subject research at the University of Minnesota granted approval for the study.

### Stimuli

The stimuli consisted of recordings of 72 natural sounds and 18 piano notes. The natural sounds included human vocal sounds (both speech and non-speech, e.g. baby cry, laughter), animal cries (e.g. monkey, lion), musical instruments (e.g. piano, flute), and tool sounds (e.g. scissors, vacuum cleaner). The piano notes were sampled from an online database (University of Iowa Musical Instrument Samples; [www.theremin.music.uiowa.edu](http://www.theremin.music.uiowa.edu); accessed on 11 July 2012), and consisted of six musical notes (i.e. C, D, E, G<sup>b</sup>, A<sup>b</sup>, and B<sup>b</sup>) sampled from three octaves (middle octave, with C at 254 Hz, and the two octaves directly above). Sounds were sampled at 16 kHz and their duration was cut at 1000 ms. Sound onset and offset were ramped with a 10 ms linear slope, and their energy (RMS) levels were equalized. Inside the scanner - before starting the measurement - sounds were played to the subject while earplugs were in place. Intensity of the sounds was further adjusted in order to equalize their perceived loudness. Sounds were presented using the fMRI-compatible S14 model earphones of Sensimetrics Corporation ([www.sens.com](http://www.sens.com)), with a linear frequency transfer up to 8 kHz.

## Magnetic Resonance Imaging

Data were acquired on a 7T whole body system driven by a Siemens console using a head gradient insert operating at up to 80 mT/m with a slew rate of 333 T/m/s. A head RF coil (single transmit, 16 receive channels) was used to acquire anatomical ( $T_1$  weighted) and functional ( $T_2^*$  weighted BOLD) images.  $T_1$  weighted ( $1\text{ mm}^3$ ) images were acquired using a modified Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) sequence (TR=2500 ms; TI=1500 ms; TE=3.67 ms). Proton density (PD) images were acquired together with the  $T_1$  weighted images (both acquisitions are interleaved in the modified MPRAGE sequence) and were used to minimize inhomogeneities in  $T_1$  weighted images (van de Moortele et al., 2009). Acquisition time for anatomy was ~ 7 minutes.  $T_2^*$  weighted functional data were acquired using a clustered Echo Planar Imaging (EPI) technique.

The experiment was designed according to a fast event-related scheme. The acquisition parameters were: TR = 2600 ms; time of acquisition [TA] = 1200 ms; TE = 30 ms; number of slices = 31; GRAPPA acceleration X3; partial Fourier 6/8; voxel size =  $1.5 \times 1.5 \times 1.5\text{ mm}^3$ , silent gap = 1400 ms. Slices covered the brain transversally from the inferior portion of the anterior temporal pole to the superior portion of the superior temporal gyrus bilaterally. The experiment consisted of six runs. In three runs (run 1/3/5) the 72 natural sounds were presented 3 times in total (i.e. each sound was presented in one time per run). In the three remaining runs (run 2/4/6) the 18 different piano notes were presented 12 times in total (4 times per run). For the notes only, sound presentation was clustered according to octave. That is, each run consisted of three blocks, and per block all sounds belonging to one of the octaves were presented in succession.

Within each run, sounds were randomly spaced at a jittered interstimulus interval of 2, 3, or 4 TRs and presented – with additional random jitter - in the silent gap between acquisitions. Zero trials (trials where no sound was presented 5% of the trials), and catch trials (trials in which the sound which was just heard was presented, 8% and 5% of the trials in ‘natural sounds’ and ‘notes’ runs respectively) were included. Subjects were instructed to perform a one-back task, and were required to respond with a button press when a sound was repeated. Catch trials were excluded from the analysis. Each run lasted approximately 10 minutes.

Functional and anatomical data were analyzed with BrainVoyager QX. Pre-processing consisted of slice scan-time correction (with sinc interpolation), 3-dimensional motion correction, temporal high and low pass filtering (removing drifts of 7 cycles or less per run, and smoothing with 2 time points), and spatial

smoothing (2 mm). Functional data were co-registered to the anatomical data and normalized in Talairach space (Talairach and Tournoux, 1988). Functional data were resampled (with sinc interpolation) in the normalized space at a resolution of 1 mm isotropic. Anatomical volumes were also used to derive gray matter segmentations indicating the border between white and gray matter. Using this border, inflated hemispheres of the individual subjects were obtained.

### Computing spectral tuning of cortical locations

Based on the responses to natural sounds, we calculated the voxels' spectral profile using customized Matlab code ([www.mathworks.com](http://www.mathworks.com)). We followed methodological procedures similar to those previously described in Moerel et al. (2012) and De Martino et al. (2013).

In this analysis, the natural sounds are passed through a biologically plausible computational model of auditory processing from the cochlea to the mid-brain (Chi et al., 2005), characterizing them according to their frequency content  $\mathbf{W}$  ( $N \times F$ ; where  $N$  = number of sounds and  $F$  = number of frequency bins). We calculated the voxels spectral preference (matrix  $\mathbf{R}$  [ $F \times V$ ], where  $V$  = number of voxels) by combining the sounds frequency representation  $\mathbf{W}$  with the fMRI response matrix  $\mathbf{Y}$  [ $N \times V$ ]. That is, for each voxel  $j$  its frequency profile  $R_j$  [ $(F \times 1)$ ] was obtained as the relevance vector machine (RVM) solution to the linear problem:

$$Y_j = \mathbf{W} \cdot R_j$$

where each element  $i$  of the vector  $R_j$  describes the contribution of the frequency bin  $i$  to the overall response of voxel  $j$  (for details see Moerel et al., 2012). We obtained a tonotopic map by considering the characteristic frequency (CF) of a voxel as the maximum of the coefficients in  $R_j$ . Tonotopic cortical maps for each subject were obtained by logarithmic mapping of best-frequency values to a red-yellow-green-blue colour scale.

Group maps were created by first optimizing alignment across subjects using functional cortex based alignment (fCBA). FCBA complements standard CBA (Goebel et al., 2006; Frost and Goebel, 2012) by allowing an individually defined functional region, in addition to the major sulci and gyri, to drive the across-subject alignment. We based fCBA on the tonotopic maps, by delineating the low frequency region consistently present in the vicinity of Heschl's gyrus as region of interest. The resulting alignment information was used for calculating and displaying group cluster maps (see below). For comparison, group maps were additionally created by averaging single-subject maps in Talairach space.

### Identification of octave-tuned locations

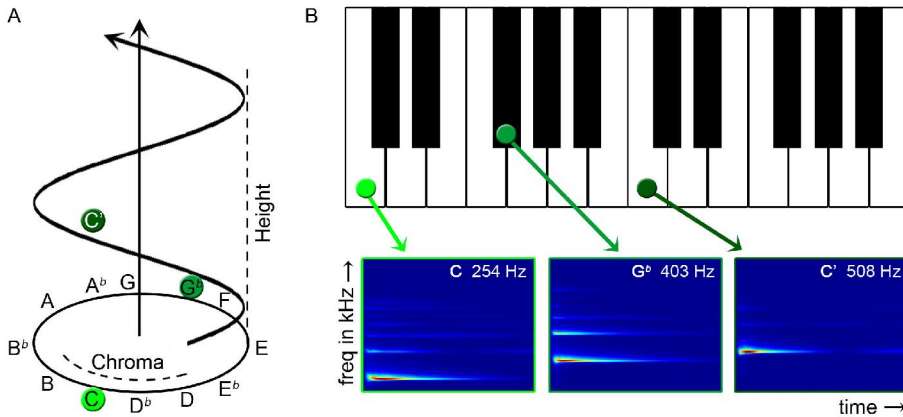
In order to characterize the complex, multi-peaked spectral tuning of voxels as estimated based on their responses to natural sounds (i.e. voxels' spectral tuning beyond their main frequency peak), we performed a clustering analysis dividing the voxels into groups of similar spectral tuning (see also Moerel et al., submitted; see chapter 4 for details). Specifically, we computed the normalized autocorrelation function of each profile  $R_i$ , computed the correlation between the autocorrelation profiles of the voxels, and obtained a fully data-driven subdivision of this network into spatially non-overlapping clusters with the Louvain algorithm (<http://www.brain-connectivity-toolbox.net>; Blondel et al., 2008; Rubinov and Sporns, 2011).

We computed the characteristic profile (i.e. the centroid) of each resulting cluster by averaging the normalized spectral profiles of the voxels in that cluster. Cluster spatial maps were obtained by computing the correlation of each voxels' autocorrelation profile to the corresponding clusters' centroid, reflecting the representativeness of that voxel for the cluster. Group maps were created by transforming individual maps into fCBA space, averaging matching maps across subjects, and smoothing resulting maps on the group surface (repeat value = 1).

### Assessing octave invariance based on responses to notes

The structure of musical pitch can be modelled as a helix, where the circular dimension reflects changes in pitch chroma repeating every octave and the vertical dimension reflects changes in pitch height (see Figure 1A; Shepard et al., 1982; Warren et al., 2003). Responses to different musical notes in voxels whose activity reflects the representation of notes in octave invariant manner should follow the circular chroma dimension. As such, similarity in responses to two notes should decrease linearly as the frequency distance between them grows. Similarity should be minimal at a frequency distance of  $\frac{1}{2}$  octave, yet as the frequency distance between notes continues to grow, their distance in octaves decreases. Responses to notes exactly one octave apart should be most similar again ('octave-invariant dissimilarity hypothesis'; see Figure 2A).

To investigate whether octave-tuned voxels display an octave-invariant response pattern, we performed the two separate analyses. First, we assessed octave-invariance for predicted responses to musical notes (prediction based on responses to natural sounds only; data of run 1/3/5). Considering predicted responses to notes explores whether octave-based neuronal tuning would suffice to represent notes in octave-invariant manner. Second, we assessed oc-



**Figure 1. Structure of musical notes**

(A) The pitch helix, with a circular chroma dimension and vertical height dimension, reflects the structure of musical pitch. (B) Examples of piano notes used in the experiment: the lower C with main frequency peak at 254 Hz (light green); the lower Gb with main frequency peak at 403 Hz (middle green); the middle C with main frequency peak at 508 Hz (dark green).

tave-invariance for measured responses to musical notes (based on the data of run 2/4/6). By investigating the measured responses to musical notes, we determine whether the auditory cortex truly operates in this fashion.

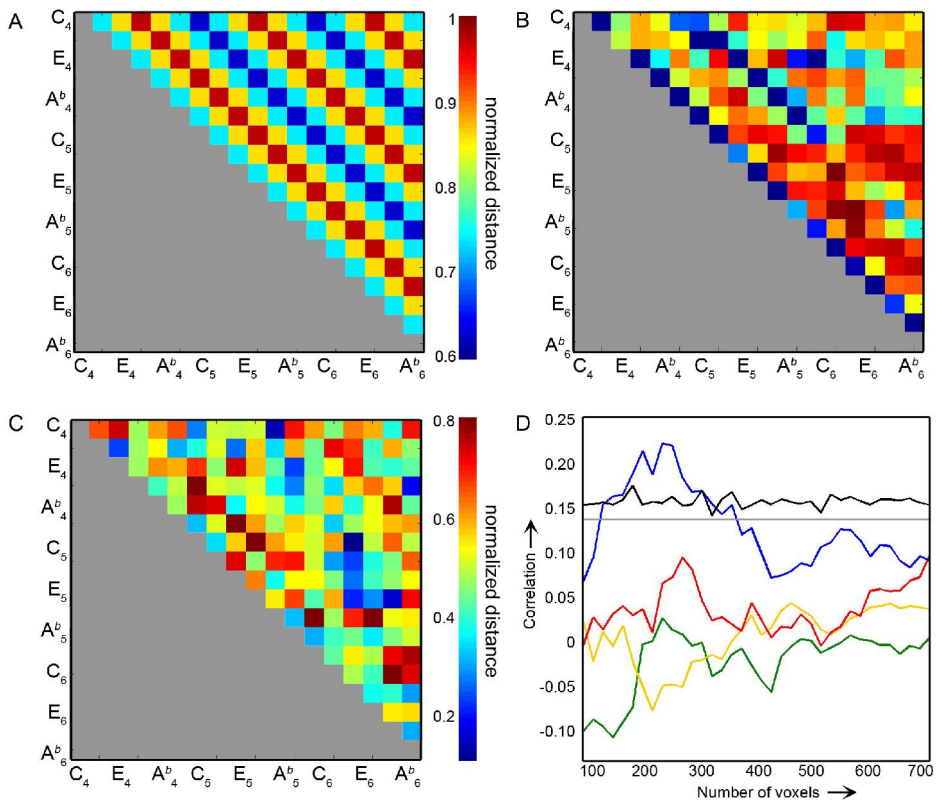
We investigated whether predicted responses to musical notes display an octave-invariant response pattern in the following manner. First, we predicted how voxels across the auditory cortex (independent of their cluster membership) respond to musical notes. That is, we used the spectral profile of each voxel  $j$  to predict the response  $\hat{Y}_j$  to the 18 notes as

$$\hat{Y}_j = \mathbf{W} \cdot \mathbf{R}_j$$

where  $\mathbf{W}$  [ $N \times F$ ] is the frequency representation of the musical notes ( $N = 18$ ). Second, for each cluster  $c$ , we selected the voxels belonging to that cluster  $V_c$ . For each note  $N_k$ , we identified those voxels  $V_{c,k}$  whose main tonotopic peak coincided with the main frequency of  $N_k \pm 0.1$  octave. We selected a varying number of voxels per note  $N_k$ , based on the correlation of the voxels' spectral profile to the centroid of cluster  $c$  (range of 90-720 voxels selected, reflecting a selection of 5-40 voxels per note).

Third, we built distance matrix  $\mathbf{D}$  – reflecting the dissimilarity (Kriegeskorte et al., 2008) in brain responses to musical notes – per index. That is, for each combination of notes  $N_{k,l}$  we computed  $D_{k,l}$  as  $(1 - \text{linear correlation between } \hat{Y}_k \text{ and } \hat{Y}_l)$ . Here,  $\hat{Y}_k$  and  $\hat{Y}_l$  reflect the predicted response to sounds  $N_k$  and  $N_l$  respectively, in the combined set of voxels  $V_{c,k}$  and  $V_{c,l}$  concatenated across the six subjects (fixed effects group analysis).





**Figure 2. Similarity of notes in octave invariant locations**

(A) The hypothesized distance matrix across responses in octave invariant locations: responses are similar to sounds at a lag of one octave, and sounds at non-octave frequency intervals elicit different patterns of activation. (B) The normalized distance matrix in predicted responses to notes, based on spectral profiles estimated on the natural sounds and the frequency content of musical notes (252 frequency matched voxels; correlation with hypothesized pattern  $r = 0.49$ ,  $p < 0.01$ ). (C) The normalized distance matrix of observed responses to notes in octave-invariant voxels (252 frequency matched voxels; correlation with hypothesized pattern  $r = 0.22$ ;  $p < 0.01$ ). (D) Octave invariance (correlation to the hypothesized distance matrix) per cluster, across a varying number of voxels (selected by their correlation to their centroid). Green, yellow, red, and blue lines correspond to clusters A-D in Figure 3 respectively (i.e. the octave cluster is shown in blue). Grey and black lines indicate the  $p < 0.05$  threshold based on parametric testing and permutation testing respectively (see Methods).

We conducted this analysis at group level because only a limited number of voxels met the selection requirements for each note at the level of single subjects. Finally, we computed the linear correlation between distance matrix **D** and the octave-invariant dissimilarity matrix to assess octave-invariance in the voxels' representation. By computing the linear correlation of subsets of **D** and the

octave-invariant dissimilarity matrix, we assessed octave-invariance for notes at one octave distance (similarity between the first and second, and the second and third octave together). Furthermore, we assessed octave-invariance for all octave combinations separately, thus for the one octave lag between the first and second octave, the one octave lag between the second and third octave, and the lag of two octaves between the first and third octave.

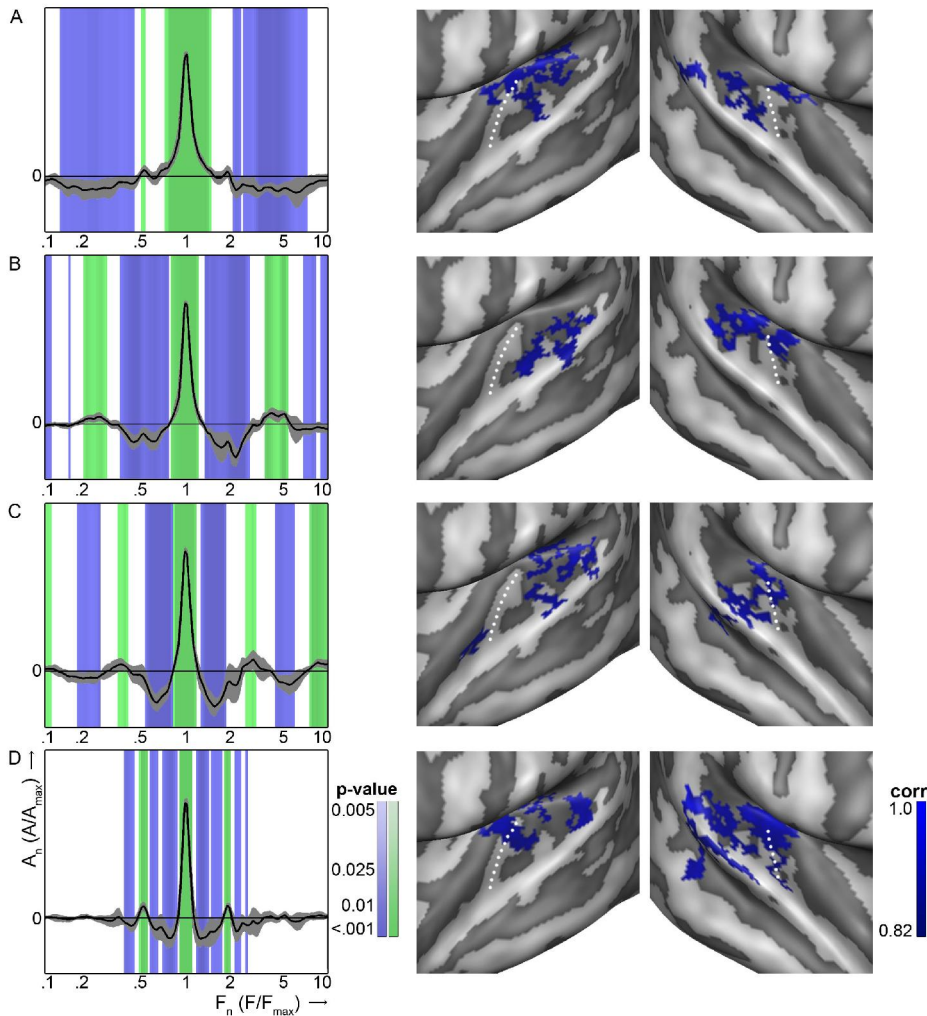
After analyzing predicted responses  $\hat{\mathbf{Y}}$  to test that octave-tuned voxels could reflect musical notes in octave-invariant manner, we analyzed the responses measured during the presentation of musical notes (data from runs 2/4/6). To this end, we computed the cortical responses to the piano notes with a single subject General Linear Model (GLM) analysis (Friston et al., 1995) creating matrix  $\mathbf{Y}$ . Following analyses were exactly the same as described above. Statistical significance was assessed at group level, by comparing the obtained correlation between  $\mathbf{D}$  and the octave-invariant dissimilarity matrix to the null-distribution obtained by randomly permuting  $\mathbf{D}$  ( $N = 1000$  permutations).

## Results

### Validation of spectral tuning clusters

In accordance with our previous results, we identified different types of spectral tuning in human auditory cortex (Figure 3). That is, we replicated four of the five spectral tuning types as identified in our previous study (Moerel et al., submitted). Each of these clusters contained an approximately equal number of voxels (23.6, 19.8, 21.1, and 32.6 percent of total number of voxels included in clusters 1-4 respectively). Specifically, we found a broadly tuned cluster, a cluster with an additional peak around  $4 \cdot F_{\max}$ , and a cluster tuned to  $3 \cdot F_{\max}$  (Figure 3A-C respectively). As expected, voxels in the fourth spectral cluster were tuned to multiple frequency bands at one octave interval from each other (i.e. the centroid displayed an additional peak at  $2 \cdot F_{\max}$ ; Figure 3D).

The cortical location of each cluster at group level is shown on the right side of Figure 3. The first cluster (Figure 3A) stretched along Heschl's sulcus (HS) and the medial extremity of Heschl's gyrus (HG) bilaterally, occupying posterior PAC in the group maps. The second cluster (Figure 3B) occupied parts of planum temporale (PT) and part of middle superior temporal gyrus (STG). The third cluster (Figure 3C) was located on HG in the right hemisphere and on PT



**Figure 3. Clusters' centroids and maps**

Centroids (left) and group maps (right) characterizing the four extracted clusters. The centroids' main peak represents the main frequency peak across voxels, and additional peaks show the presence of sensitivity to additional frequency bands at consistent spectral intervals. Positive and negative significant deviations from zero are colour coded in green and blue respectively. Grey areas show the standard error in extracted centroids across the six subjects. Corresponding maps were extracted per subject, projected to the cortex, and averaged in fCBA space. (A-D) show the broadly tuned cluster (green in Fig. 2C); the cluster tuned to  $4 \cdot F_{max}$  (yellow in Fig. 2C); the cluster tuned to  $3 \cdot F_{max}$  (red in Fig. 2C); and the octave cluster tuned to  $2 \cdot F_{max}$  (blue in Fig. 2C). The colors show the correlation of each voxel in the map to its corresponding centroid; the white dotted line indicates the location of HG.

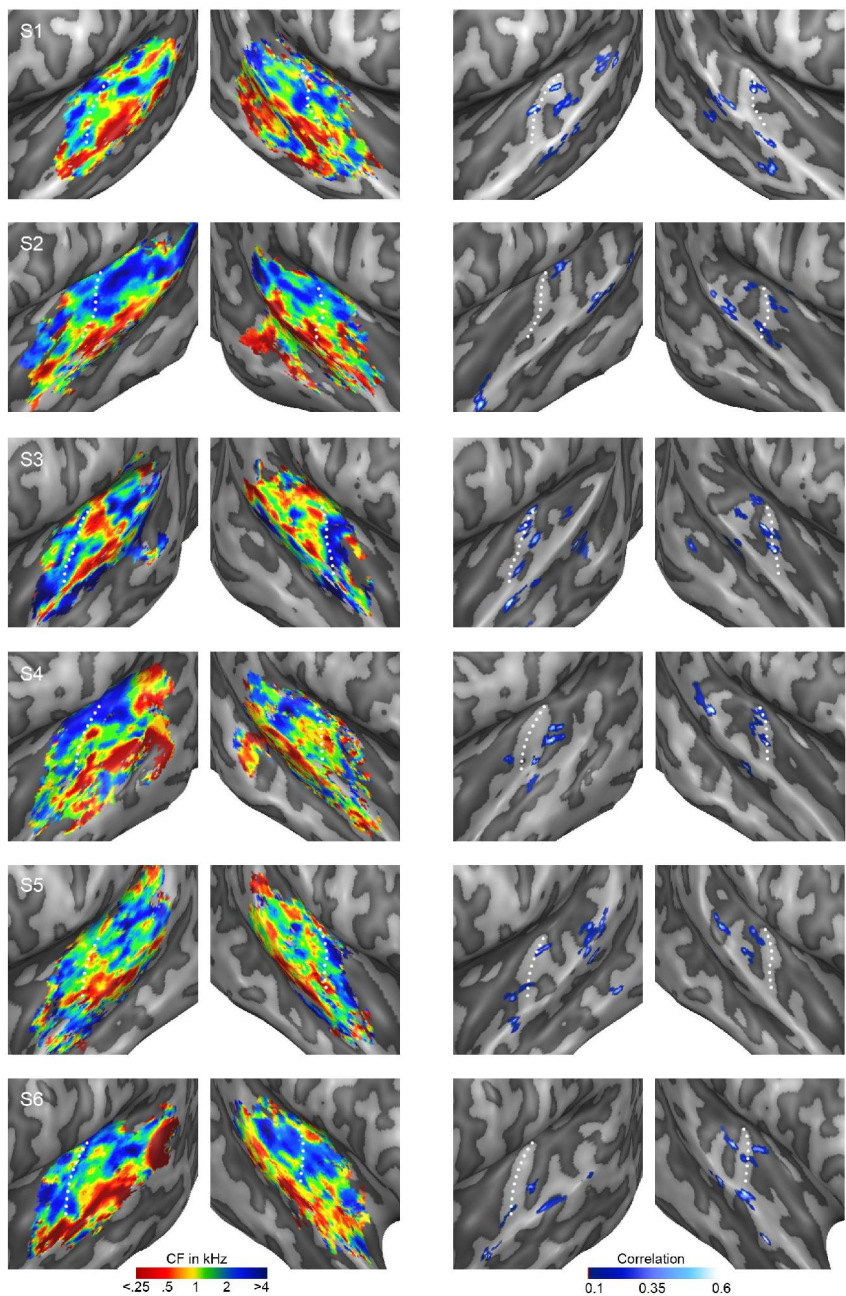
in the left hemisphere. Finally, the fourth octave-tuned cluster (Figure 3D) extended from HG posteriorly onto PT and along the STG.

### Invariance in octave-tuned locations

We assessed octave-invariance for predicted responses to musical notes (prediction based on responses to natural sounds only; half of the data) and for measured responses to musical notes (other half of the data; see Methods). Specifically, we computed the distance in responses to musical notes, and compared this distance to that expected if an octave-invariant representation would be in place (i.e. lowest distance for notes 1 octave apart; highest distance for notes  $\frac{1}{2}$  octave apart; see Figure 2A).

Based on predicted responses to musical notes, octave tuned locations responded to notes in an octave-invariant manner (as shown by a high correlation between the predicted distance matrix and the octave-invariant dissimilarity matrix (compare Figure 2A-B; correlation  $r = 0.49$ ). Neuronal populations in the remaining three clusters displayed a substantially lower correlation with the octave-invariant dissimilarity matrix ( $r = 0.30/0.27/0.32$ , for the three clusters respectively). Interestingly, octave invariance in the octave cluster was stronger for a lag of one octave (ranges between octaves one/two, and two/three;  $r = 0.64$  and  $r = 0.54$  respectively) than a lag of two octaves (range between octaves one/three;  $r = 0.39$ ).

While considering predicted responses to notes shows that octave-based neuronal tuning would suffice to represent these notes in octave-invariant manner, investigating the measured responses to musical notes may reveal whether the auditory cortex truly operates in this fashion. Indeed, measured responses to musical notes showed a similar pattern as predicted responses. Figure 2D shows the octave-invariance for a lag of one octave (invariance between first and second octave, and invariance between second and third octave combined) in measured responses to notes per cluster and for a varying number of voxels. Specifically, the correlation between the measured distance matrix and the octave-invariant dissimilarity matrix is shown. Octave invariance was largest in the octave-tuned cluster, and peaked at a cortical size of 252 voxels. Around this peak only, statistical testing revealed a significant octave-invariance at group level (permutation testing;  $p < 0.01$ ). The distance matrix across all three octaves is shown in Figure 2C. As in the predicted responses, the octave invariance was stronger for a lag of one octave (ranges between octaves one/two, and two/three;  $r = 0.22$ ,  $p < 0.05$  and  $r = 0.25$ ,  $p < 0.05$  respectively) than a lag of two octaves (range between octaves one and three;  $r = -0.15$ ).

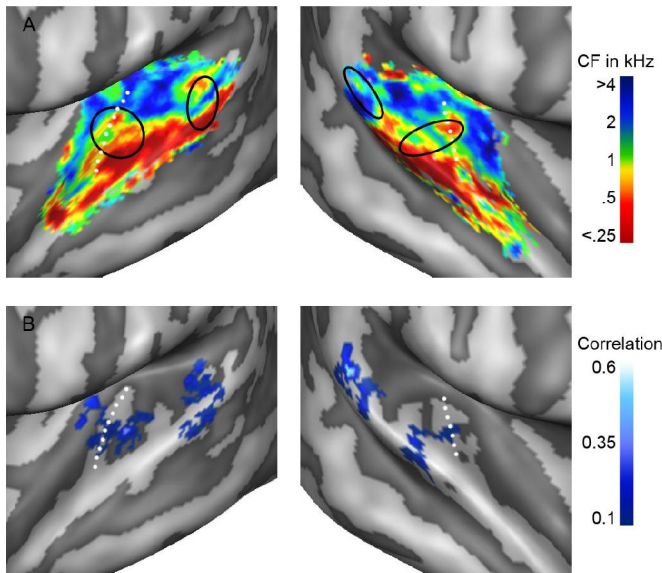


**Figure 4. Individual tonotopy maps and octave invariant locations**

Maps of tonotopy are shown in the left column. In each subject and hemisphere, a high-low-high frequency gradient reflecting primary regions, and additional frequency clusters are present throughout the superior temporal plane. Octave invariant maps are shown in the right, and display the correlation of voxel profiles to the octave-tuned centroid. The white dotted line indicates the location of HG.

### Location of octave-invariant voxels

Next, we examined the location of the octave-invariant voxels (i.e. the 252 frequency-matched voxels most characteristic of the octave cluster). At the level of single subjects (see Figure 4), these octave-invariant voxels clustered on HG and at its posterior adjacency in HS. Beyond the HG region, voxels clustered on PT and posterior STG. Group maps, obtained after aligning subjects in fCBA space, replicated this pattern. Specifically, the voxels clustered most densely on HG and just posterior to it in HS (Figure 5B). An additional octave invariant region was present on posterior STG bilaterally. Group maps based on averaged single-subjects maps aligned in Talairach space showed a very similar - although altogether more spread - pattern (maps not included).



**Figure 5. Group tonotopy and 'octave invariance' map**

Group tonotopy map (top), and group octave invariant locations (bottom), created by averaging single subject maps in fCBA space. The octave invariant cluster occupied locations on HG, HS, and posterior STG bilaterally, which are outlined on the tonotopy maps by black circles. Octave invariant maps display the correlation of voxel profiles to the octave-tuned centroid averaged across subjects. The white dotted line indicates HG.

In order to further explore the location of the octave-invariant cluster, we compared it – at single subject and group level – to extracted tonotopic maps. Consistent with results of previous studies, tonotopic maps displayed a large low frequency region near HG, bordered anteriorly (on HG and first transverse sulcus [FTS]) and posteriorly (on HS and anterior PT) by regions preferring higher frequencies (Da Costa et al., 2011; Formisano et al., 2003; Humphries et al., 2010; Moerel et al., 2012). Beyond Heschl's region, on PP, PT, and STG/STS additional frequency clusters were present (see Figure 4 and 5A and single

subject and group maps respectively). Bilaterally, the octave-invariant cluster occupied a region preferring relatively low frequencies in the centre of HG. The additional octave-invariant regions on posterior STG occupied regions of both low and high frequency preference (black outline in Figure 5A).

## Discussion

In this study, we used ultra-high field functional MRI to investigate the representation of musical notes in the human auditory cortex. First, we replicated our earlier results, by showing that a subset of cortical populations displayed spectral tuning to multiple frequency bands at exactly one octave lag. Second, we showed that these populations contained an octave-invariant representation of piano notes, as they responded similarly to notes one octave apart, and differently to notes at other musical lags. The octave-invariant cortical locations were clustered most densely on the centre of HG, in a region that preferred relatively low frequencies. We suggest that this region reflects human primary auditory cortex (PAC), and consequently propose that finely tuned neuronal populations in human PAC underlie the human octave percept.

### Octave-based spectral tuning in human auditory cortex

Our observation of cortical populations tuned to multiple frequencies at exactly one octave is in accordance with electrophysiological recordings in various mammals, showing sensitivity to two tones at octave lag sequentially presented (Brosch et al., 1999; Brosch & Schreiner, 2010) or neuronal receptive fields with multiple octave-related peaks (Kadia & Wang, 2003; Norena et al., 2008). Our data does not allow inferring which type of spectral tuning at the level of individual neurons underlies our observations. We measured responses in voxels that contain thousands of neurons, at a temporal resolution in the range of seconds. Consequently, we may have observed combination-sensitive mechanisms operating at the level of single neurons (Sadagopan and Wang, 2009; Wang et al., 2005), but the results could - with equal likelihood - reflect the complex average spectral profile of many simple, single-peaked neurons. To advance our understanding regarding the neuronal tuning underlying our octave-tuned clusters, a well-controlled exploration of non-linear and combination-sensitive tuning to the multiple frequency bands within these profiles is needed.



### Octave invariant percept as early as primary auditory cortex

Hierarchical models of auditory perception suggest that lower processing levels – such as the PAC – are limited to low-level acoustic feature analysis (Rauschecker et al., 1995; Wessinger et al., 1998). Processing complexity is proposed to increase when moving towards anterior and lateral locations (e.g. anterior and middle STG; Rauschecker and Scott, 2009), and in these regions a more abstract and perceptual sound representation may emerge (Belin et al., 2000; Binder et al., 2000; Warren et al., 2003). In the current study, we observed an octave invariant representation on HG. Based on cytoarchitectonic studies (Galaburda and Sanides, 1980; Morosan et al., 2001) and our observation of preference for low frequencies this region (Da Costa et al., 2011; Formisano et al., 2003; Moerel et al., 2012), we interpret this location as the border between primary regions hA1/hR, possibly extending to the human homologue of middle lateral belt regions on HS (border between hML/hAL). In this region, the sound representation followed the human percept rather than the linear increase in the sounds' frequency content. Consequently, our results add to the accumulating evidence that the PAC, the earliest auditory cortical region, is much more than a simple sensory feature detector, as it contributes to complex and relatively abstract sound representations. This conclusion conflicts with a strictly hierarchical view of auditory processing. However, it is in line with several more recent findings showing for example that 25-50% of A1 neurons in the awake mammal are unresponsive to simple tones and instead respond only to a complex combination of sound features (Sadagopan and Wang, 2009); that feature tuning in A1 neurons is strongly modulated by context and task demands (David et al., 2012; Fritz et al., 2003); and that early cortical levels are involved in creating an abstract representation of speech sound (Formisano et al., 2008). Together, these results suggest that a viewing PAC as a simple feature detector is insufficient to explain processing in this region and that within PAC sound representation of at least intermediate complexity resides.

Beyond PAC, octave invariant locations observed in the current study traced a pathway through HS (presumably belt) to posterior STG (presumably parabelt). Based on the relative stability in responses to stimuli changed in pitch height but stable in chroma observed in these regions, we suggest that a pathway extending posteriorly from PAC preserves a representation of pitch chroma. This conclusion is in apparent conflict with a recent study by Warren and colleagues (2003), who concluded that pitch chroma was processed anterior to PAC, while pitch height was processed in posterior regions. These opposing conclusions may result from the radically different experimental and analysis approaches used in the two studies. Specifically, Warren et al. (2003) based



their conclusion on observations of stronger responses to changes in pitch chroma and pitch height anterior and posterior to PAC respectively. We report similarity in response patterns across changes in chroma and height, rather than the mean activation level reported by Warren et al. (2003). Furthermore, in our study pitch chroma changes were always accompanied by pitch height changes, making a direct comparison across studies impossible. Further studies are needed to unravel a mechanism that supports both findings.

### **Octave equivalence at two octaves**

Based on predicted responses to notes, we showed that the multi-peaked spectral tuning in the auditory cortex is more apt in representing invariance at one octave lag than at a lag of two octaves. This reduction of octave invariance with increasing distance between notes may reflect human perception. Surprisingly few studies investigated octave equivalence outside a musical setting and beyond a range of one octave (Ward, 1954). Moreover, these studies produced conflicting results (Hoeschele et al., 2012). Further explorations of human octave perception are needed to explore whether our results may directly reflect the subjects' percept. In the measured responses to musical notes, we did not observe octave invariance at a lag of two octaves, possibly due to insufficient power in our experimental setup. As our fMRI measurements were inevitably noisy, the invariant pattern at a lag of two octaves may have been too small to pick up.

### **Functional relevance of octave invariance**

The octave-tuned neuronal populations observed in the current study - clustered most densely on HG/HS and posterior STG - may be at the basis of the universal phenomenon of octave perception. As such, these populations may reflect the biological basis of 'internal octave templates', as proposed in psycho-physiological models (Demany and Semal, 1988). In addition to their relation to musical perception, populations with octave-based spectral tuning may support the analysis of vocalizations. Vocalizations - including speech, human voices, and animal sounds - are periodically structured (Ross et al., 2007; Norena et al., 2008), and typically have most energy in the fundamental and first harmonic spectral peaks (Schwartz et al., 2003). Octave-based tuning could enhance neural activity when the fundamental and the first harmonic of a complex stimulus are simultaneously present (Kadia and Wang, 2003), and thereby enhance the most relevant part of an incoming vocalization. A similar mechanism might serve the processing of vowels specifically, as an octave ratio between the first

and second formant of vowels is relatively common (Schwartz et al., 2003). Again, octave-based tuning may serve to enhance those spectral components that are most relevant for the recognition of vowels, and suppress noise.

Note that the existence of a biological substrate for octave perception does not imply that this phenomenon is inborn. In fact, it is entirely feasible that these finely tuned neuronal populations, and accordingly the octave percept, emerge with learning early in life. Consequently, it would be interesting to evaluate whether similarly tuned neuronal populations can be observed in young children and monkeys. Furthermore, exploring brain plasticity by investigating musicians is of interest, as such studies could reveal whether with training other musical intervals may become equally hard-wired in the brain.

## References

- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403: 309-312.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10: 512-528.
- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech* P10008.
- Brosch M, Schreiner CE (2000) Sequence sensitivity of neurons in cat primary auditory cortex. *Cerebral Cortex* 10: 1155-1167.
- Brosch M, Schulz A, Scheich H (1999) Processing of sound sequences in macaque auditory cortex: Response enhancement. *J Neurophysiol* 82: 1542-1559.
- Chi T, Ru P, Shamma SA (2005) Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am* 118: 887-906.
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci* 31: 12067-14075.
- David SV, Fritz JB, Shamma SA (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc Natl Acad Sci U S A* 109: 2144-2149.
- De Martino F., Moerel M., van de Moortele PF, Ugurbil K, Yacoub E, Formisano E (2013) Spatial organization of sound frequency preference and selectivity in the human inferior colliculus. *Nat Comm* 22, doi: 10.1038/ncomms2379.
- Demany L, Armand F (1984) The perceptual reality of tone chroma in early infancy. *J Acoust Soc Am* 76: 57-66.
- Demany L, Semal C (1988) Dichotic fusion of two tones one octave apart: Evidence for internal octave templates. *J Acoust Soc Am* 83: 687-695.
- Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science* 322: 970-973.
- Formisano E, Kim D, Di Salle F, van de Moortele PF, Ugurbil K, et al. (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40: 859-869.
- Friston KJ, Frith CD, Turner R, Frackowiak RS (1995) Characterizing evoked hemodynamics with fMRI. *Neuroimage* 2: 157-165.
- Fritz JB, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience* 6: 1216-1223.
- Frost M, Goebel R (2011) Measuring structural-functional correspondence: Spatial variability of specialized brain regions after macro-anatomical alignment. *Neuroimage* 59:1369-1381.
- Galaburda A, Sanides F (1980) Cytoarchitectonic organization of the human auditory cortex. *The Journal of Comparative Neurology* 190: 597-610.
- Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp* 27: 392-401.
- Hoeschele M, Weisman RG, Sturdy CB (2012) Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. *Atten Percept Psychophys*, DOI:10.3758/s13414-012-0364-2.
- Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. *Neuroimage* 50: 1202-1211.
- Kadia SC, Wang X (2003) Spectral integration in A1 of awake primates: Neurons with single- and multi-peaked tuning characteristics. *J Neurophysiol* 89: 1603-1622.
- Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. *Nature* 452: 352-356.
- King AJ, Nelken I (2009) Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nature Neuroscience* 12: 698-701.

- Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis – connecting the branches of systems neuroscience. *Front Syst Neurosci* 2. doi: 10.3389/neuro.06.004.2008.
- Merzenich MM, Brugge JF (1973) Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res* 50: 275-296.
- Merzenich MM, Knight PL, Roth GL (1973) Representation of cochlea within primary auditory cortex in the cat. *Brain Res* 63: 343-346.
- Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci* 32: 14205-14216.
- Moerel M, De Martino F, Santoro R, Ugurbil K, Goebel R, Yacoub E, Formisano E (submitted) Processing of natural sounds: Characterization of multi-peak spectral tuning in human auditory cortex.
- Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, et al. (2001) Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13: 684-701.
- Norena A, Gourevitch B, Pienkowski M, Shaw G, Eggermont JJ (2008) Increasing spectrotemporal sound density reveals an octave-based organization in cat primary auditory cortex. *J Neurosci* 28: 8885-8896.
- Randel, D. M., (ed.). *The Harvard Dictionary of Music* (2003). Cambridge Massachusetts, Harvard University Press (4th ed.).
- Rauschecker JP (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111-114.
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12: 718-724.
- Ross D, Choi J, Purves D (2007) Musical intervals in speech. *Proc Natl Acad Sci U S A* 104: 9852-9857.
- Rubinov M, Sporns O (2011) Weight-conserving characterization of complex functional brain networks. *Neuroimage* 56: 2068-2079.
- Sadagopan S, Wang X (2009) Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. *J Neurosci* 29: 11192-11202.
- Schwartz DA, Howe CQ, Purves D (2003) The statistical structure of human speech sounds predicts musical universals. *J Neurosci* 23: 7160-7168.
- Shepard RN (1982) Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89: 305-333.
- Talairach J, Tournoux P (1988) *Co-planar stereotaxic atlas of the human brain*. Stuttgart: G. Thieme.
- van De Moortele PF, Auerbach EJ, Olman C, Yacoub E, Ugurbil K, Moeller S (2009) T1 weighted brain images at 7 Tesla unbiased for Proton Density, T2 contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. *NeuroImage* 46: 432-446.
- Wang X (2007) A sharper view from the top. *Nature Neuroscience* 10:1509-1511.
- Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli *Nature* 435: 341-346.
- Ward WD (1954) Subjective musical pitch. *J Acoust Soc Am* 26, 369-380.
- Warren JD, Uppenkamp S, Patterson RD, Griffiths TD (2003) Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci U S A* 100: 10038-10042.
- Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (1998) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 13: 1-7.
- Wright AA, Rivera JJ, Hulse SH, Melissa S, Neiwirth JJ (2000). Music perception and octave generalization in rhesus monkeys. *J Exp Psychology* 129: 291-307.



# **Chapter 6**

## **General discussion**



The studies discussed in this thesis addressed two main research questions. First, how is the spectral content of natural sounds represented in our brain? Second, what mechanisms underlie the transition of a tonotopic sound image into a more abstract, behaviourally relevant sound representation? These questions were investigated by examining neuronal population frequency tuning based on fMRI responses to natural sounds. While each specific result was discussed in the separate chapters, the present discussion aims at integrating the findings across chapters, and at providing a more general framework. Specifically, a parcellation of the human auditory cortex based on its spectral tuning properties is proposed, and neuronal mechanisms underlying the formation of sound categories based on sensory sound representations are suggested.

## **Parcellating the human auditory cortex**

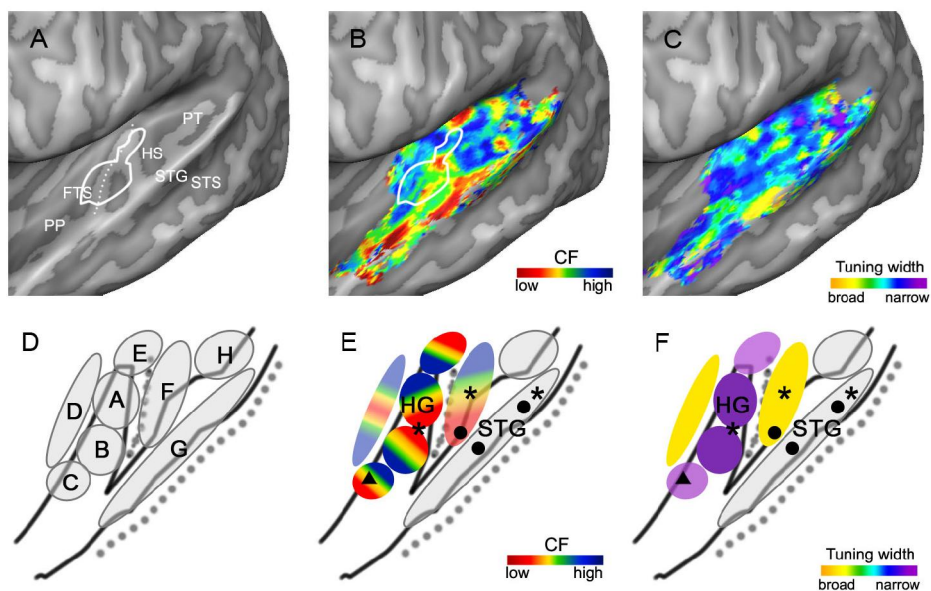
The non-invasive extraction of cortical tonotopic maps has become increasingly feasible in recent years, yet the interpretation of these maps is still under debate (Da Costa et al., 2011; Langers and van Dijk, 2012). Below, tonotopic maps are combined with tuning width maps (Moerel et al., 2012) and anatomical information (De Martino et al., submitted; Dick et al., 2012; Morosan et al., 2001; 2005) to interpret the frequency gradients that cover the human superior temporal plane.

### **Primary auditory cortex**

Reversals in tonotopic gradients can segregate primary cortical regions from each other. However, tonotopic gradients continue at the boundary between primary and non-primary auditory cortex, and consequently this border can only be defined based on features other than tonotopy. In chapter 2, core primary regions were distinguished from the adjacent belt cortex by using tuning width maps (i.e. the selectivity of a neuronal population response to the preferred frequency) in addition to best frequency maps (Moerel et al., 2012; Rauschecker and Tian, 2004; Rauschecker et al., 1995). Specifically, a narrowly tuned region was observed along HG bilaterally. This finding was replicated at 7T in chapter 3 (De Martino et al., 2013; see also Figure 1A-C for result in single subject at 7T). When interpreting only the narrowly tuned portion of the tonotopy map as the PAC (Brodmann area 41; Brodmann, 1909), a mirror-symmetric tonotopic gradient was objectively distinguished running in posterior-medial to anterior-lateral direction along HG (covering regions 'A' and 'B' in Figure 1D). Although this orientation of the human PAC does not coincide with interpreta-



tions of recent tonotopy studies (Da Costa et al., 2011; Humphries et al. 2010), it is in agreement with previous anatomical and functional descriptions (Dick et al., 2012; Formisano et al., 2003; Galaburda and Sanides, 1980; Morosan et al., 2001). Regions ‘A’ and ‘B’ in Figure 1D may reflect the human homologues of monkey primary regions A1 and R respectively (hA1 and hR), each including a full tonotopic gradient. In accordance with cyto- and myeloarchitectonic studies (Hackett et al., 1998; 2001), the region defined as hA1 (region ‘A’ in Figure 1) coincides with the location of highest myelin-related contrast in this subject, as measured in a related study in which high resolution anatomical images were obtained in addition to tonotopic maps (De Martino et al., submitted). Note that the octave-invariant locations observed in chapter 5 clustered most densely in the low frequency region between hA1 and hR (“stars” in Figure 1E-F; see below).



**Figure 1. Tonotopy and tuning width in a single subject at 7T**  
The top row shows (A) anatomy, with the white dotted line indicating the location of HG, (B) tonotopy, and (C) tuning width in the left hemisphere of a single subject. In panels A and B, the narrowly tuned region on HG is outlined in white. In the bottom row, these functional maps are related to cytoarchitectonic regions (see main text). Specifically, on HG a narrow core (purple) is surrounded by more broadly tuned regions (yellow; panel F). This core contains a mirror-symmetric tonotopic map, and additional frequency gradients are present at the medial and lateral extremities of HG (panel E). Circles, stars, and triangles indicate regions where voice-sensitive, octave-invariant, and harmonically tuned locations were observed respectively. The empty circles in panels E-F indicate regions where topographic maps are present, but are not yet interpreted.

Region 'C' in Figure 1 contains another full tonotopic gradient, and – together with its anatomical location – this region is interpreted as reflecting the human homologue of primary region RT (hRT). The tuning width of this region is narrow (see chapters 2-3), but it is less heavily myelinated than regions hA1 and hR (De Martino et al., submitted; Dick et al., 2012). This corresponds with studies investigating the monkey auditory cortex, which showed that RT is the least primary-like of the primary cortical regions (Hackett et al., 1998). Note that the harmonically tuned neuronal populations observed in chapter 4 clustered most densely in hRT (triangles in Figure 1E-F; see below).

### Non-primary regions

The tonotopic maps shown in chapters 2-5 and in Figure 1 extend well beyond the primary areas. Although interpretation beyond the primary regions remains difficult, some topographic gradients emerged consistently across subjects and experiments. Within the medial belt region indicated with 'D' in Figure 1D, a high-low-high frequency gradient resides. This gradient was not clear in the group 3T maps of chapter 2, but was repeatedly observed in the 7T maps where a higher spatial resolution was available (both at individual subject and group level; chapters 4-5). Region 'D' may reflect two fields of the human medial belt cortex (Brodmann area 52; Brodmann, 1909), each containing a full tonotopic gradient. These regions could correspond to the human homologues of regions RM and MM in the monkey (Kusmieriek and Rauschecker, 2009). Furthermore, another reversal in frequency is present medial to the PAC (region 'E' in Figure 1D). That is, this region contains a low-to-high frequency gradient in postero-medial to antero-lateral direction. Region 'E' was narrowly tuned and more heavily myelinated than other non-primary areas. Accordingly, it may be the most primary-like belt region, and could correspond to monkey regions CM/CL, which were reported to display primary-like characteristics (Hackett et al., 1998).

Immediately posterior to the PAC (regions hA1 and hR), the lateral belt is located (region 'F' in Figure 1D). This region is distinguished from the PAC by its broader tuning width. The region contains a single high-to-low tonotopic gradient in postero-medial to antero-lateral direction (following the tonotopic gradient in hA1). Although only one tonotopic gradient occupies this region, it may comprise of two functionally separate subfields. The medial part of this region, including a full high-to-low tonotopic gradient, is part of the 'octave-tuned pathway' described in chapter 5. Conversely, the lateral part of this region, tuned to low frequencies only, coincides with the voice region on middle STG (chapter 2). These two subfields within region 'F' could correspond to the human homo-

logues of lateral belt regions AL and ML in the monkey (Hackett et al., 1998). Note that while ML in the monkey contains a full tonotopic gradient (Rauschecker and Tian, 2004), the complete hypothesized hML is tuned to low frequencies. This could reflect a specialization of this region unique to humans, following the need to process the low spectral energy of speech specifically.

Beyond these regions, additional gradients of tonotopy and tuning width occupied regions on the STG and the posterior end of the temporal plane ('G' and 'H' in Figure 1D). Based on their location, they most likely reflect parabelt regions (Galaburda and Sanides, 1980; Hackett et al., 1998; Morosan et al., 2005). Within these regions, voice sensitive and octave representing voxels reside (indicated by circles and stars respectively in Figure 1E-F; see chapters 2 and 5 respectively). Understanding the precise topology of these regions requires additional knowledge regarding their response properties (e.g. tuning to temporal/spectral modulations, latency).

## **Neuronal mechanisms underlying sound categorization**

In humans, sound processing is commonly described either in terms of low level tuning to a range of acoustic features (e.g. Da Costa et al., 2011; Schönwiesner and Zatorre, 2009), or in terms of higher-level regions that contain an abstract sound representation (partially) independent of acoustic features (e.g. Belin et al., 2000; Leaver et al., 2010). This section integrates the findings of the current thesis, and proposes that fine-grained spectral sensitivity of neuronal populations underlies the abstraction of sound representations beyond their acoustic features.

### **Fine spectral tuning links sensory to higher-level sound representations**

Three findings of this thesis suggest that spectral tuning mechanisms in the human auditory cortex are functionally and behaviourally relevant. First, similar to the visual cortex (Hasson et al., 2003; Rajimehr et al., 2011), in the auditory cortex an intrinsic relation exists between category-sensitive regions and low-level topographic maps (see chapter 2). Second, neuronal populations throughout the auditory cortex exhibit sensitivity to multiple frequency bands, which may be shaped by the acoustic environment (selective tuning to harmonic and octave frequency lags; chapter 4). Third, within octave-tuned neuronal populations only, an octave-invariant sound representation resides (chapter 5).

These fMRI observations may reflect neuronal filtering mechanisms operating at early or intermediate stages of the transformation from tonotopic sound

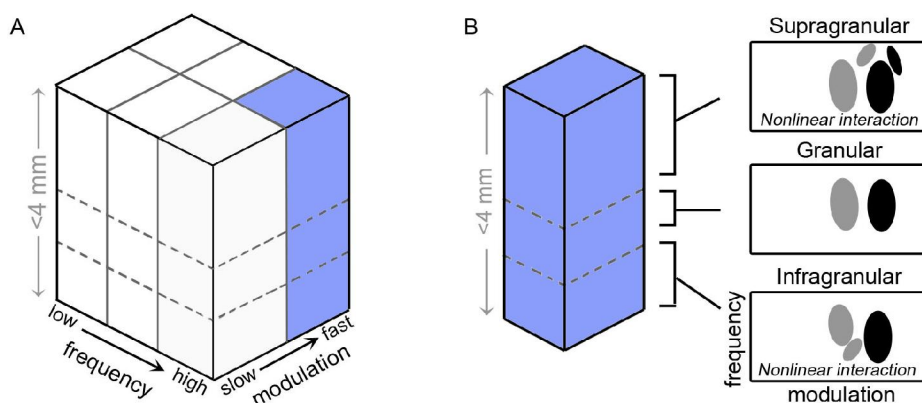
images into their representations in an abstract and categorical manner. Such mechanisms could function by enhancing relevant spectral content and filtering out noise. In the case of tuning to low frequencies in voice-sensitive regions (observed in chapter 2), this mechanism could serve to selectively amplify the spectral content that is crucial for the processing of human vocalizations. Similarly, the harmonically tuned neuronal populations (observed in chapter 4) could serve to significantly enhance harmonic components relative to background noise without harmonic structure (Kadia and Wang, 2003).

Furthermore, multi-peaked neuronal populations (observed in chapters 4-5) could help categorizing sounds. For example, the octave-tuned voxels responded in similar manner to tones with frequencies at a 2:1 ratio (see chapter 5) and thereby effectively categorized sounds according to their pitch chroma (e.g. similar responses across voxels to 'C of octave 1' and 'C of octave 2'). Finally, the multi-peaked tuning could serve as a feature detection mechanism. Multi-peaked neurons observed in monkey cortex based on electrophysiological recordings were often observed to be combination-sensitive. That is, neurons responded strongly to a specific mixture of sound features while each component alone failed to elicit an excitatory response (Wang et al., 1995; Sadagopan & Wang, 2009). Neurons with combination-sensitive spectral tuning only respond if a specific combination of frequency bands is present in the sounds. Such nonlinear neuronal tuning, finely tuned to the acoustic environment, could serve as feature detectors for processing a wide range of natural sounds (deCharms et al., 1998; Fitzpatrick et al., 1993; Woolley et al., 2005; Wang et al., 1995).

### **Future directions**

Future studies are needed to explore where and when in the brain these neuronal filtering mechanisms emerge. The current thesis showed multi-peaked and perceptually relevant sensitivity to multiple frequency bands as early as PAC (octave-tuned populations; see chapter 6). Although this is in contrast with purely hierarchical models of auditory processing that generally view PAC as a simple frequency processor (Rauschecker and Scott, 2009), this finding is in accordance with recent observations in various animal models (Brosch and Schreiner; Wang). In fact, it was suggested that the construction of a more abstract, invariant representation might be progressively implemented throughout PAC cortical layers (Atencio et al., 2009; Sharpee et al., 2011). Within primary auditory regions, a precise microcircuit is implemented. Orthogonal to the cortical sheet, neuronal feature tuning may be relatively stable (Atencio and Schreiner, 2009; Schreiner, 2009). This vertical arrangement throughout the

thickness of the cortical sheet is referred to as a 'columnar' organization (see Figure 2A). While neuronal features are relatively stable within a column (Phillips & Irvine, 1981), tuning is not identical. Instead, feature tuning becomes more nonlinear with increased distance from granular layers (Atencio et al., 2009; Sadagopan & Wang, 2009; Sharpee et al., 2011; see Figure 2B). Accordingly, the sound representation may become more variation tolerant, abstract, and categorical in superficial layers of PAC (Atencio et al., 2009), ultimately forming a representation of sounds that is no longer a faithful replica of its acoustic structure. By combining an individual parcelation of the auditory cortex (see Figure 1) with advances in ultra- high field fMRI, it may be possible to study the transformation of spectral sound representations within functionally distinct auditory subfields. By exploiting technical advances, the spatial resolution of fMRI measurements can be pushed to obtain a reliable signal from different cortical layers (Polimeni et al., 2010; Yacoub et al., 2007). This may allow exploring receptive field nonlinearities throughout the layers of cortex and evaluating the sound representation per cortical layer.



**Figure 2. Hypothesized columnar and laminar organization in PAC**

(A) Orthogonal to the cortical sheet of PAC, a columnar organization is hypothesized to reside, referring to a relative stable feature preference in vertical direction. For example, the blue bar prefers high frequencies and fast temporal modulations. (B) Feature tuning within a column is similar but not identical, as illustrated by the receptive fields of granular input layers versus infra- and supragranular output layers on the right. Simple tuning to features in granular input layers are nonlinearly combined across layers, resulting in complex tuning in infra- and supragranular output layers.

An additional challenge for future studies will be to investigate when these fine-grained spectral tuning mechanisms arise. Are these complex tuning properties innate? How do they emerge with normal development? Are they plastic at later stages in life? Recent studies emphasize the flexibility of feature preference, for

example with training (Ohl and Scheich, 1997) but also at the short temporal scale of contextual changes or task instruction (David et al., 2012; Fritz et al., 2003). The exploration of these outstanding questions may be tackled using the methods proposed and validated in this thesis.

## References

- Atencio CA, Schreiner CE (2009) Laminar diversity of dynamic sound processing in cat primary auditory cortex. *J Neurophysiol* 103: 192-205.
- Atencio CA, Sharpee TO, Schreiner CE (2009) Cooperative nonlinearities in auditory cortical neurons. *Neuron* 58: 956-966.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000). Voice-selective areas in human auditory cortex. *Nature* 403: 309-312.
- Brodmann K (1909) Vergleichende lokalisationslehre der Grosshirnrinde. Leipzig: JA Barth.
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RSJ, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci* 31: 12067-14075.
- De Martino F, Moerel M, van de Moortel PF, Ugurbil K, Goebel R, Yacoub E, Formisano E (2013) Spatial organization of frequency preference and selectivity in the human inferior colliculus. *Nat Comm* 22, doi: 10.1038/ncomms2379.
- De Martino F, Moerel M, Xu J, van de Moortele PF, Ugurbil K, Goebel R, Yacoub E, Formisano E (submitted) Using functional and anatomical contrast at 7 Tesla to localize primary auditory regions in vivo.
- David SV, Fritz JB, Shamma SA (2012) Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc Natl Acad Sci U S A* 109: 2144-2149.
- deCharms RC, Blake DT, Merzenich MM (2004) Optimizing sound features for cortical neurons. *Science* 280: 1439-1443.
- Dick F, Taylor Tierney A, Lutti A, Josephs O, Sereno MI, Weiskopf N (2012) In vivo functional and myeloarchitectonic mapping of human primary auditory areas. *J Neurosci* 32: 16095-16105.
- Fitzpatrick DC, Kanwal JS, Butman JA, Suga N (1993) Combination-sensitive neurons in the primary auditory cortex of the mustached bat. *J Neurosci* 13: 931-940.
- Formisano E, Kim D, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40: 859-869.
- Fritz JB, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience* 6: 1216-1223.
- Galaburda AM, Sanides F (1980) Cytoarchitectonic organization of the human auditory cortex. *The Journal of Comparative Neurology* 190: 597-610.
- Hackett TA, Preuss TM, Kaas JH (2001) Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* 441: 197-222.
- Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394: 475-495.
- Hasson U, Harel M, Levy I, Malach R (2003) Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* 37: 1027-1041.
- Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. *Neuroimage* 50: 1202-1211.
- Kadia SC, Wang X (2003) Spectral integration in A1 of awake primates: Neurons with single- and multi-peaked tuning characteristics. *J Neurophysiol* 89: 1603-1622.
- Kusmirek P, Rauschecker JP (2009) Functional specialization of medial auditory belt cortex in the alert rhesus monkey. *J Neurophysiol* 102: 1606-1622.
- Langers DRM, van Dijk P (2012) Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cerebral Cortex* 22: 2024-2038.
- Leaver AM, Rauschecker JP (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J Neurosci* 30: 7604-7612.
- Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci* 32: 14205-14216.
- Moerel M, De Martino F, Santoro R, Ugurbil K, Yacoub E, Formisano E (submitted) Processing of natural sounds: Characterization of multi-peak spectral tuning in human auditory cortex.

- Morosan P, Rademacher J, Schleicher A, Amunts A, Schormann T, Zilles K (2001) Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13: 684-701.
- Morosan P, Schleicher A, Amunts K, Zilles K (2005) Multimodal architectonic mapping of human superior temporal gyrus. *Anat Embryol* 210: 401-406.
- Ohl FW, Scheich H (1997) Learning-induced dynamic receptive field changes in primary auditory cortex of the unanaesthetized Mongolian gerbil. *J Comp Physiol* 181: 685-696.
- Phillips DP, Irvine DRF (1981) Responses of single neurons in physiologically defined primary auditory cortex (AI) of the cat: frequency tuning and responses to latency. *J Neurophysiol* 45: 48-58.
- Polimeni JR, Fischl B, Greve DN, Wald LL (2010) Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *Neuroimage* 52: 1334-1346.
- Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RBH (2011) The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biology* 9: doi:10.1371/journal.pbio.1000608.
- Rauschecker JP, Tian B (2004) Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91: 2578-2589.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111-114.
- Sadagopan S, Wang X (2009) Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. *J Neurosci* 29: 11192-11202.
- Schonwiesner M, Zatorre RJ (2009) Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proc Natl Acad Sci U S A* 106: 14611-14616.
- Schreiner CE (2009) Spectral processing in auditory cortex. In: *The Auditory Cortex* (Winer JA and Schreiner CE, eds), pp275-308. New York: Springer.
- Sharpee TO, Atencio CA, Schreiner CE (2011) Hierarchical representations in the auditory cortex. *Current Opinion in Neurobiology* 21: 1-7.
- Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74: 2685-2706.
- Woolley SMN, Fremouw TE, Hsu A, Theunissen FE (2005) Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience* 8: 1371-1379.
- Yacoub E, Shmuel A, Logothetis N, Uğurbil K. (2007) Robust detection of ocular dominance columns in humans using Hahn Spin Echo BOLD functional MRI at 7 Tesla. *Neuroimage* 37: 1161-1177.





## Summary

In everyday life, we are constantly surrounded by sounds. Within the various structures that compose the human auditory system, acoustic features of the incoming sounds are processed by specialized neural mechanisms and transformed into behaviourally useful representations. Here, two research questions were addressed. First, how is the frequency content of natural sounds represented in the human brain? Second, what mechanisms underlie the transition of a frequency-based (i.e. tonotopic) sound image into a more abstract, behaviourally relevant representation?

Chapter two investigated spectral tuning in the human auditory cortex by combining fMRI measurements during natural sound stimulation with mathematical modelling of the brain responses. This novel experimental approach allowed defining topographic maps of frequency preference (tonotopy) and selectivity (tuning width). The combination of these maps defined a parcellation of the auditory cortex into functional subfields. In contrast with observations in earlier studies, results indicated that the topographic representation of frequency extends well beyond the primary regions and also covers higher-order and category-selective auditory regions. Specifically, regions with preferential responses to human voice occupy the low frequency portions of the tonotopic map. This low frequency bias in voice regions likely reflects the selective amplification of relevant and category-characteristic spectral bands, serving as a useful processing step for transforming sensory (tonotopic) images of vocal sounds into higher-level neuronal representations.

Chapter three exploited the advantages of ultra-high field imaging (7 Tesla) to map the spatial organization of spectral responses in a small human midbrain structure crucial for auditory processing: the human inferior colliculus (hIC). For the first time in human subjects, one tonotopic high-to-low gradient was observed within the hIC. This frequency gradient is oriented in dorso-lateral to ventro-medial direction. Conversely, the width of the spectral tuning is organized in radial manner, and varies from narrow to broad when moving from the middle to the extremity of isofrequency contours. Finally, tuning width is narrower in hIC than on the auditory cortex. These findings are very relevant for future studies of audition, as they pave the way for studying the interplay between human subcortical and cortical neuronal populations non-invasively.

While in chapter two the characteristic frequency preference of neuronal populations was examined, the fourth chapter investigated cortical spectral tuning beyond its main tonotopic peak by means of a data-driven clustering analy-

sis. Beyond neuronal populations with single-peaked frequency tuning, approximately 60% of auditory populations displayed sensitivity to multiple frequency bands. Specifically, neuronal populations sensitive to multiple frequency bands (I) at exactly one octave distance from each other, (II) at harmonically related frequency intervals, and (III) with no apparent relationship were observed. It was suggested that beyond the well-known cortical tonotopic organization, multi-peaked spectral tuning might amplify selected combinations of frequency bands. Such selective amplification could serve to detect complex sound features and aid in segregating sounds of interest in noisy auditory scenes. Furthermore, these results put forward the hypothesis that prominent perceptual phenomena such as the perceptual similarity of tones and melodies that differ by one or multiple octaves (octave invariance) can be explained by these tuning properties.

Chapter five specifically tested the hypothesis that octave-based multi-peaked spectral tuning - as observed in chapter four - underlies the perceptual phenomenon of octave invariance. Using the same methods as in chapter four, octave-tuned neuronal populations were identified on primary auditory cortex (PAC) and the superior temporal gyrus (STG). In these octave-tuned locations, but not other auditory cortical locations, predicted as well as measured responses to musical notes confirmed the hypothesis of octave-based representations of sound frequency. These findings indicate that neuronal populations at the earliest cortical stage of auditory processing (PAC) played an active role in the formation of higher-level sound representations, beyond the sensory mapping of frequency.

In summary, the series of fMRI studies in this thesis examined the spectral representation of natural sounds in human subcortical and cortical auditory regions, describing in detail the large-scale topographic maps of frequency preference (i.e. tonotopy) and selectivity (i.e. tuning width). The studies also revealed that beyond their main tonotopic peak, neuronal populations in human auditory cortex are sensitive to behaviourally relevant combinations of multiple frequency bands. This complex tuning is hypothesized to reflect neuronal filtering mechanisms operating to transform tonotopic sound images into their higher-level representations. Future studies are needed to explore how, where and when these neuronal filtering mechanisms emerge.

## Samenvatting

In ons dagelijkse leven worden we continu omringd door geluiden. Deze binnenkomende geluiden worden binnen de verschillende hersenstructuren van het menselijke auditieve systeem ontleed in hun akoestische kenmerken. Vervolgens worden de geluiden gerepresenteerd op een manier die nuttig is voor ons gedrag. In dit proefschrift worden twee onderzoeksvragen besproken. Ten eerste, hoe worden frequenties van natuurlijke geluiden gerepresenteerd in het menselijke brein? Ten tweede, welke mechanismen zorgen voor de transformatie van een tonotopische geluidsrepresentatie naar een abstracte geluidsrepresentatie, die relevant is voor menselijk gedrag?

Door het wiskundig modelleren van fMRI reacties op natuurlijke geluiden, onderzocht hoofdstuk twee verwerking van frequenties in de menselijke auditieve cortex. Met behulp van deze nieuwe methode werden topografische kaarten van frequentie voorkeur (tonotopie) en frequentie selectiviteit gevonden. Door deze kaarten te combineren, kon de auditieve cortex in verschillende velden worden verdeeld. In tegenstelling tot wat in eerdere studies gevonden werd, bleken deze kaarten tot ver buiten de primaire auditieve gebieden te reiken. De kaarten bedekken ook hogere gebieden in de auditieve hiërarchie, inclusief gebieden die gevoelig zijn voor auditieve categorieën. Gebieden gevoelig voor mensenstemmen liggen in stukken van de tonotopische kaart die afgestemd zijn op lage frequenties. Deze voorkeur voor lage frequenties in stemgevoelige gebieden zou een selectieve versterking van relevante en categoriekarakteristieke frequenties kunnen reflecteren. Zulke versterking zou als eerste stap kunnen dienen in de verandering van een sensorische (tonotopische) geluidsrepresentatie naar een neuronale representatie van hoger niveau.

In hoofdstuk drie werden de voordelen van ultrahoog veld imaging (7 Tesla) gebruikt om de ruimtelijke organisatie van frequentie verwerking in een kleine auditieve structuur van de menselijke hersenen te onderzoeken: de menselijke inferieure colliculus (mIC). Voor de eerste keer in mensen, werd een tonotopische gradiënt gezien binnen de mIC, die van hoog naar laag loopt in dorso-laterale naar ventro-mediale richting. In tegenstelling tot frequentie voorkeur, is frequentie selectiviteit radiaal georganiseerd. Frequentie selectiviteit varieert van scherp in het midden van iso-frequentie contouren, naar breed aan de rand van deze contouren. Als laatste bleek dat frequentie selectiviteit scherper is in mIC dan in de auditieve cortex. Deze bevindingen zijn relevant voor toekomstige auditieve studies, omdat ze een weg banen voor het non-invasief bestuderen van de interactie tussen menselijke subcorticale en corticale gebieden.

Terwijl in hoofdstuk twee de karakteristieke frequentie voorkeur van neuronale populaties werd onderzocht, onderzocht het vierde hoofdstuk frequentie verwerking buiten deze tonotopische voorkeur aan de hand van een data-gedreven clusteranalyse. Naast neuronale populaties met voorkeur voor één frequentie, bleek dat ongeveer 60% van de auditieve cortex voorkeur heeft voor meerdere frequenties. Neuronale populaties met voorkeur voor meerdere frequenties (I) die precies een octaaf bij elkaar vandaan liggen, (II) met harmonische afstand tot elkaar, en (III) met onduidelijke relatie tot elkaar werden gevonden. Deze resultaten suggereren dat naast de bekende corticale tonotopische organisatie, complexe afstemming op meerdere frequenties zorgt voor versterking van geselecteerde combinaties van frequenties. Zulke selectieve versterking kan zorgen voor detectie van complexe geluidskenmerken en kan helpen in het opsplitsen van auditieve omgevingen. Verder ondersteunen deze resultaten de hypothese dat prominente perceptuele fenomenen, zoals de perceptuele gelijkheid tussen tonen en melodieën die een of meerdere octaven van elkaar verschillen (octaaf invariantie), verklaard kunnen worden aan de hand van deze complexe afstemming op frequenties.

Het vijfde hoofdstuk testte de hypothese dat de afstemming op octaven die in hoofdstuk vier gevonden werd, ten grondslag ligt aan het perceptuele fenomeen van octaaf invariantie. Aan de hand van de methoden gebruikt in hoofdstuk vier, werden neuronale populaties afgestemd op octaven geïdentificeerd in de primaire auditieve cortex (PAC) en de superieure temporale gyrus (STG). Op deze plekken, maar niet op andere corticale locaties, bevestigden zowel voorspelde als gemeten reacties op muzikale noten de octave-invariante frequentie-representatie hypothese. Deze bevindingen laten zien dat neuronale populaties het vroegste corticale niveau in auditieve verwerking (PAC), naast het representeren van frequentie, een actieve rol spelen in het vormen van een abstracte geluidsrepresentatie.

Samengevat onderzocht deze reeks fMRI studies de representatie van frequenties in natuurlijke geluiden in menselijke subcorticale en corticale auditieve gebieden. De studies beschreven omvangrijke topografische kaarten van frequentie voorkeur (tonotopy) en frequentie selectiviteit in detail. Verder lieten de studies zien dat buiten hun tonotopische voorkeur, neuronale populaties in de menselijke hersenen gevoelig zijn voor meerdere frequenties wiens combinaties relevant zijn voor menselijk gedrag. Deze complexe frequentie voorkeur kan een neurale filter mechanisme reflecteren, dat ervoor zorgt dat tonotopische geluidsbeelden getransformeerd worden in een representatie van hoger niveau. Toekomstige studies zijn nodig om te onderzoeken hoe, waar, en wanneer deze neurale filter mechanismen ontstaan.

## Acknowledgements

First and foremost, I want to thank my promotor Elia Formisano. Elia, your enthusiasm and energy were paramount in the creation of this dissertation. You have taught me to solve problems, to keep asking questions, and never give up (even after many rejections). Thank you for giving me the opportunity to work with such independence, while making time for me whenever it was needed.

Many thanks also to Rainer Goebel, my co-promotor. Rainer, our 'BrainMatters' project gave me a chance to work with you. Your dedication and optimism are an inspiration to me, and with me to the entire Cognitive Neuroscience group.

I owe a lot of gratitude to the people who I had the pleasure of collaborating with me over the last years. Roberta, your critical opinion during all our discussions never failed to deepen my understanding of our ongoing projects. I really enjoyed working together with you. Fabrizio, your help saved me when I was still working on my master thesis (although you maybe didn't realize it at the time), and you have been supportive ever since. Giancarlo, your advice and patience prevented me from making a lot of mistakes in my analyses. I would like to thank the members of the auditory group: Milene, Lars, Nick, Anke, Lars, and Kiki, for their help and fruitful suggestions over the last years. Many thanks also to Essa Yacoub in Minneapolis, for making it possible to explore our auditory hypotheses at the best spatial resolution.

Annemie, Christl en Riny, zonder jullie zou dit proefschrift meerdere malen gestrand zijn. Bedankt voor jullie organisatie en geduld.

I'm very grateful for the friendship of my colleagues. In particular, my long-term study friends Tom and Anke, and my office mates Job and Marieke. Tom, thank you for being my friend for all these years. I enjoy talking to you so much, because your outspoken opinions make me see things in a different way. You and Job together are simply amazing. The two of you make me laugh every time, and your friendship means the world to me. Job, you are the sweetest person I know. I can't thank you enough for your support on days when getting a dissertation together wasn't so much fun, and joining me in celebrating on the days that it was. Anke, you are without doubt my favorite person to drink wine together with. Thank you for all the fun we have together, and for all the serious conversations too. I still hope to keep you around longer! Marieke, you made

sure I was never alone in the office and beat me every single time at staying at uni late. Thank you for our conversations, your experience gives me perspective on what's to come.

Martin, Nick, Alard, Kamil, Aline, Marin, Britta, Roberta, Lars, Gesa, Felix, Matteo, Joao, Rosanne, Alex, Anna, Martin, Jan, and Anna, thanks for all the fun lunches, relaxing picknicks, and great dinners. You make it great to come to work every day.

Ook mijn vriendinnen buiten mijn werk wil ik hier bedanken, voor de welkome afleiding en gezelligheid tijdens de afgelopen jaren. Ingrid en Aimée, jullie ziekenhuis-verhalen zorgden er de laatste vier jaar voor dat ik vertrouwen hield in mijn eigen beroepskeuze J. Ook al zien we elkaar al lang niet meer elke dag, jullie vriendschap is heel belangrijk voor me. Ik weet dat ik altijd op jullie kan rekenen en dat waardeer ik enorm. Loes, Gwenny, Susanne, Simone, Lilli, Carlijn, Leonie en Yoanne, bedankt voor alle gezellige avonden en zondagse lunches! Ik hoop dat we snel weer samen zijn in Maastricht, Utrecht, Hilversum, of ergens ertussen in. Simone, jou wil ik speciaal bedanken voor je vriendschap. Ik kijk er altijd naar uit om met je te kletsen, natuurlijk onder het genot van wijn en kaasjes!

Lieve pap en mam, ik kan jullie niet genoeg bedanken voor jullie liefde, goede zorgen en jullie absolute vertrouwen in mij. Pap, je geeft me het gevoel dat wat er ook gebeurt, ik bij jou terecht kan. Jij bent mijn stabiele basis. Mam, jouw enthousiasme en optimisme laten me zien dat alles mogelijk is, en dat je van het leven moet genieten. Door jullie samen voel ik de vrijheid om dingen te proberen, grenzen te verleggen en te genieten van elke dag. Ik ontspan volledig zodra ik bij jullie thuis kom en ik vind het altijd heerlijk om bij jullie te zijn.

Denise, je bent mijn liefste zusje. Jouw kracht en onafhankelijkheid zijn een voorbeeld voor mij en ik kijk enorm naar je op. Ik ben trots op je, kijk er elke keer naar uit om je te zien, en ben heel blij dat onze band zo goed is.

Federico, you have challenged me since the day we met. You make me grow both in my work and as a person, and I'm deeply grateful for that. At the same time you read me so well that you manage to understand perfectly when I need your support instead of a challenge. Thank you for your unfailing support when I need it. You make me think, laugh, and experience, and I couldn't ask for anything more than that.

## Curriculum vitae

Michelle Maria Lucia Moerel was born on August 15<sup>th</sup> 1985 in Geleen (the Netherlands). In 2003 she graduated from high school at the Graaf Huyn College in Geleen, and she subsequently enrolled at the Faculty of Psychology and Neuroscience at Maastricht University. In her third year, she studied a semester abroad at Bond University, Australia. Michelle obtained her Bachelor of Science, cum laude, in Biological Psychology in 2006. Next, she followed the Research Master program in Cognitive Neuroscience at Maastricht University. In her Master thesis, under supervision of Professor Francesco di Salle, she conducted empirical work investigating frequency tuning in the human auditory cortex. She completed the Research Master program successfully in 2008 (cum laude). Funded by an NWO TopTalent grant, she started her Ph.D. program under supervision of Professor Elia Formisano at the Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University. Here, she currently continues her work as a postdoctoral fellow.





### Articles

*\*Authors share equal contribution*

De Martino F\*, **Moerel M**\*, van de Moortele PF, Ugurbil K, Goebel R, Yacoub E, Formisano E (2013) Spatial organization of frequency preference and selectivity in the human inferior colliculus. *Nat Comm* 22: doi 10.1038/ncomms2379.

**Moerel M**, De Martino F, Formisano E (2012) Processing of Natural Sounds in Human Auditory Cortex: tonotopy, spectral tuning and relation to voice-sensitivity. *J Neurosci* 32: 14205-14216.

De Martino F, Schmitter S, **Moerel M**, Tian J, Ugurbil K, Formisano E, Yacoub E, van de Moortele PF (2012) Spin Echo functional MRI in bilateral auditory cortices at 7 Tesla: an application of B1 shimming. *NeuroImage* 63: 1313-1320.

Sack AT, Cohen Kadosh R, Schuhmann T, **Moerel M**, Walsh V, Goebel R (2009) Optimizing functional accuracy of TMS in Cognitive Studies: A comparison of methods. *J Cogn Neurosci* 21: 207-221.

**Moerel M**, De Martino F, Santoro R, Ugurbil K, Goebel R, Yacoub E, Formisano E (submitted) Processing of natural sounds: Characterization of multi-peak spectral tuning in human auditory cortex.

De Martino F, **Moerel M**, Xu J, Auerbach E, van de Moortele PF, Ugurbil K, Goebel R, Yacoub E, Formisano E (submitted) In-vivo localization of primary auditory regions using functional and anatomical contrast at 7T.

Santoro R, **Moerel M**, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E (submitted) Multiscale analysis of natural sounds in human auditory cortex.

Guerreiro MJS, Eck J, **Moerel M**, Evers L, van Gerven P (submitted) Attentional modulation in visual and auditory category-selective cortical areas.

Guerreiro MJS, Eck J, **Moerel M**, Evers L, van Gerven P (submitted) Attentional modulation of cortical activity in older age depends on sensory modality.

**Moerel M**, De Martino F, Santoro R, Ugurbil K, Yacoub E, Formisano E (in preparation) The perceptual constancy of musical notes across octaves emerges from multi-peak spectral tuning in primary auditory cortex.

De Martino F, Xu J, Auerbach E, **Moerel M**, Ugurbil K, Formisano E, Yacoub E (in preparation) Multiband GE EPI improves auditory functional responses to simple and complex sounds.

## Book Chapters

Formisano E, **Moerel M**, Bonte M (in press) Functional MRI of the auditory cortex. In: Functional MRI: From Nuclear Spins to Brain Function (Uludag K, Ugurbil K, Berliner LJ, eds). New York: Springer.

## Abstracts

**Moerel M**, De Martino F, Santoro R, Ugurbil K, Yacoub E, Formisano E (2012) Octave-based spectral tuning in human auditory cortex. 5th Conference on Auditory Cortex, Lausanne, Switzerland.

**Moerel M**, De Martino F, Santoro R, Ugurbil K, Yacoub E, Formisano E (2012) Functional networks in auditory cortex defined by a data-driven analysis of neuronal population spectral tuning. 18th meeting of the Organization for Human Brain Mapping, Beijing, China.

**Moerel M**, De Martino F, Ugurbil K, van de Moortele PF, Formisano E, Yacoub E (2011) Tonotopic map of human inferior colliculus unravelled by functional MRI at 7T. Society for Neuroscience. Washington, DC.

**Moerel M**, De Martino F, Ugurbil K, van de Moortele PF, Formisano E, Yacoub E (2011) Tonotopic map of human inferior colliculus unravelled by functional MRI at 7T. Symposium on Advances and Perspectives in Auditory Neurophysiology. Washington, DC.

**Moerel M**, De Martino F, Formisano E (2011) Sensory and categorical sound representations are intrinsically linked in human auditory cortex. 17th meeting of the Organization for Human Brain Mapping, Quebec, Canada.

**Moerel M**, De Martino F, Formisano E (2010) Multiple topographical representations of real-life sounds in human auditory cortex. 16th meeting of the Organization for Human Brain Mapping, Barcelona, Spain.

**Moerel M**, De Martino F, Formisano E (2009) Imaging the response profile of the auditory cortex to real-life sounds. 4th Conference on Auditory Cortex, Magdeburg, Germany.